

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:09 ; Search time 49.3653 Seconds  
(without alignments)  
1562.545 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISWLFILNFMEYI.....QQKKRKVDQKSVSVSTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A Geneseq\_29Jan04:\*
- 1: Geneseqp1980s:\*
  - 2: Geneseqp1990s:\*
  - 3: Geneseqp2000s:\*
  - 4: Geneseqp2001s:\*
  - 5: Geneseqp2002s:\*
  - 6: Geneseqp2003as:\*
  - 7: Geneseqp2003bs:\*
  - 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1516	100.0	273	4	Aae13151 Human ste
2	1516	100.0	273	6	Abo44414 Human ste
3	1505	99.3	272	4	Aab99220 Human thr
4	1505	99.3	272	4	Aam78328 Human pro
5	1505	99.3	272	4	Aae13168 Human ste
6	1505	99.3	272	4	Aae13150 Human ste
7	1505	99.3	272	6	Abr62112 Human clo
8	1505	99.3	272	6	Abo44432 Human ste
9	1505	99.3	272	6	Abo44413 Human ste
10	1472	97.1	265	4	Aae13163 Human sec
11	1472	97.1	265	6	Abo44427 Human sec
12	1472	97.1	292	2	Aaw85607 Secreted
13	1472	97.1	292	4	Aae13170 Human SCR
14	1472	97.1	292	5	Abp61846 Human pol
15	1472	97.1	292	6	Abr62114 Human sec
16	1472	97.1	292	6	Abo44434 Human SCR
17	1452	95.8	292	4	Aab93875 Human pro
18	1400	92.3	251	4	Aae13153 Human mat
19	1400	92.3	251	6	Abo44417 Human ste
20	1310.5	86.4	279	4	Aae13167 Mouse ste
21	1310.5	86.4	279	6	Abo44431 Mouse ste
22	1178	77.7	239	6	Abu52396 Human GPC
23	1111	73.0	195	6	Abu52398 Human GPC
24	1107	73.0	195	6	Abu52397 Human GPC
25	903	59.6	160	4	Abb11374 Human sec

26	903	59.6	160	4	Aam79312 Human pro
27	903	59.6	160	4	Aae13149 Human ste
28	903	59.6	160	6	Abo44415 Human ste
29	656	43.3	263	6	Abr62108 Secreted
30	656	43.3	263	6	Abr62115 Secreted
31	656	43.3	263	6	Abr58489 Human sec
32	656	43.3	263	7	Ado07919 Novel pro
33	648.5	42.8	243	6	Abr62110 Secreted
34	644	42.5	265	6	Abr62113 Mouse thr
35	638	42.1	229	4	Aae13162 Mouse thr
36	638	42.1	229	6	Abo44426 Human sec
37	584.5	38.6	243	6	Aae37115 Human sec
38	577.5	38.1	243	5	Abg76508 DNA encod
39	577.5	38.1	243	6	Abr62106 Secreted
40	576.5	38.0	243	6	Aae36166 Mouse SCR
41	573	37.8	250	6	Abr62101 Secreted
42	545	35.9	222	6	Abr62107 Secreted
43	540.5	35.7	229	6	Abr62102 Secreted
44	472	31.1	234	7	Abd76146 Novel hum
45	276	18.2	46	4	Aae13155 Human SCR

#### ALIGNMENTS

RESULT 1  
AAE13151  
ID AAE13151 standard; protein; 273 AA.  
XX  
AC AAE13151;  
XX  
DT 28-JAN-2002 (first entry)  
XX  
DE Human stem cell growth factor-like protein #3.  
XX  
KW Human; stem cell growth factor-like protein; antiinflammatory; nootropic;  
KW neuroprotective; vulnery; cytostatic; anticonvulsant; immunostimulant;  
KW vasotrophic; virucide; dermatological; tranquiliser; cerebroprotective;  
KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
KW supporting factor for the proliferation of stem cell.

XX Homo sapiens.  
WO200177169-A2.  
18-OCT-2001.  
05-APR-2001; 2001WO-US011208.  
05-APR-2000; 2000US-00543774.  
28-JUN-2000; 2000US-0215733P.  
09-JAN-2001; 2001US-00757562.  
05-FEB-2001; 2001US-0266614P.  
(HYSE-) HYSEQ INC.  
(KIRI ) KIRIN BEER KK.  
Tang TV, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;  
Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
WPI; 2001-657166/75.  
N-PSDB; AAD21725.  
Novel stem cell growth factor like polypeptides and polynucleotides for  
identifying modulators useful for treating diseases such as Alzheimer's

PT disease, cancer, rheumatoid arthritis, osteoporosis.  
 PS Claim 28; Page 214-215; 232pp; English.  
 XX The patent discloses novel stem cell growth factor-like proteins and polynucleotides encoding them. Proteins of the invention are also known as supporting factor for the proliferation of stem cells (SCR-1). Stem cell growth factor-like proteins are useful for supporting proliferation or survival of a stem cell or germ cell which is preferably primordial germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem cell, haematopoietic progenitor cell, pluripotent cell, or totipotent cell. The haematopoietic progenitor cell cultured using stem cell growth factor-like proteins can replace as a graft for the bone marrow transplantation or cord blood transplantation for treating a variety of diseases such as immunodeficiency syndrome, chronic granulomatous disease, duplicated immunodeficiency syndrome, agammaglobulinaemia, Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS), thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal storage diseases such as mucopolysaccharidosis, adrenal white matter degeneration, a variety of cancer and tumours. Proteins of the invention are useful for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, thrombocytopaenia, immune deficiencies and disorders such as severe combined immunodeficiency (SCID) and autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary inflammation. Sequences of the invention are also useful in gene therapy. The present sequence is stem cell growth factor-like protein from human Sequence 273 AA;

Query Match 100.0%; Score 1516; DB 4; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-111; Indels 0; Gaps 0;  
 Matches 273; Conservative 0; Mismatches 0;

Qy 1 MGHRLISWLFILFMFMEYIGSQNASRRGRMRHNPVSGQGCGCATCDYNGCLSCRP 60  
 Db 1 MGHRLISWLFILFMFMEYIGSQNASRRGRMRHNPVSGQGCGCATCDYNGCLSCRP 60  
 Qy 61 RLFFALERIGMKQIGVCLSSCSGYYGTRYPDINKCTCKACDCTCFNKNFCTCKSGFY 120  
 Db 61 RLFFALERIGMKQIGVCLSSCSGYYGTRYPDINKCTCKACDCTCFNKNFCTCKSGFY 120  
 Qy 121 LHLGKCLDNCPEGLANNHMECVSIHVCEVSEWNPSPCTKKGKCGFRGTETRVREI 180  
 Db 121 LHLGKCLDNCPEGLANNHMECVSIHVCEVSEWNPSPCTKKGKCGFRGTETRVREI 180  
 Qy 181 IOHPSAKGNLCPTTNETRCKTVQRKKCKGKGRKRRKPKKPNKGBSKEAIPDSKSLE 240  
 Db 181 IOHPSAKGNLCPTTNETRCKTVQRKKCKGKGRKRRKPKKPNKGBSKEAIPDSKSLE 240  
 Qy 241 SSK3IPEQRNKKQKQKRVQDKQKSVSVSTVH 273  
 Db 241 SSK3IPEQRNKKQKQKRVQDKQKSVSVSTVH 273

RESULT 2  
 ID ABO44414 standard; protein; 273 AA.  
 AC ABO44414;  
 XX ABO44414;  
 XX ABO44414;  
 DT 30-SEP-2003 (first entry)  
 XX Human stem cell growth factor-like protein, SCR 1 #2.  
 DE Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
 KW immunostimulant; vulnery; haematopoietic stem cell; gene therapy;  
 KW supporting factor for proliferation of stem cells; wound healing;  
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;  
 KW bone marrow transplantation; cord blood transplantation;  
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;  
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;

KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;  
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;  
 KW cartilage remodeling; bone growth; immunosuppression; human.  
 XX Homo sapiens.  
 XX OS US2003044792-A1.  
 XX 06-MAR-2003.  
 XX 28-JUN-2001; 2001US-00894912.  
 XX 28-JUN-2000; 2000US-0215733P.  
 XX 05-FEB-2001; 2001US-0266614P.  
 XX 05-APR-2001; 2001US-0282397P.  
 XX (TANG/) TANG Y T.  
 XX (LABA/) LABAT I.  
 XX (DRMA/) DRMANAC R T.  
 XX (MIZE/) MIZE N.  
 XX (NISH/) NISHIKAWA M.  
 XX (CHAO/) CHAO C.  
 XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;  
 WPI: 2003-625403/59.  
 N-PSDB; ACH04325.  
 Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.  
 Claim 23; Page 72-73; 96pp; English.  
 The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aa from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell and an implant comprising a cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell. The SCR-1 polypeptide is useful for identifying a compound that binds to the SCR-1 polypeptide and for maintaining survival of or promoting proliferation of a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic progenitor cell. The SCR-1 polypeptide is useful for promoting wound healing. The human haematopoietic stem cell or human haematopoietic progenitor cell culture using the SCR-1 polypeptide can replace as a graft for the conventional bone marrow transplantation or cord blood transplantation. The transplantation of haematopoietic stem cells can be employed as a therapy for treating diseases such as chronic granulomatous diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,

CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia  
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's  
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a Human SCR-1  
 CC protein  
 XX  
 SQ Sequence 273 AA;

Query Match 100.0%; Score 1516; DB 6; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-111;  
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKP 60  
 DB 1 MGLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKP 60

QY 61 RLFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 120  
 DB 61 RLFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 120

QY 121 LHGLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTTETRVREI 180  
 DB 121 LHGLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTTETRVREI 180

QY 181 IQHPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNKGESKEAIPDSKSL 240  
 DB 181 IQHPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNKGESKEAIPDSKSL 240

QY 241 SSKEIPEQRNKKQKKRKVDKQKSVSVSTVH 273  
 DB 241 SSKEIPEQRNKKQKKRKVDKQKSVSVSTVH 273

## RESULT 3

AAB99220  
 ID AAB99220 standard; protein; 272 AA.

AC AAB99220;

DT 07-SEP-2001 (first entry)

XX Human thrombospondin-30.

XX Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;  
 KW malignant tumour; haemopathy; HIV infection; immunological disease;  
 KW inflammation disease.

XX Homo sapiens.

XX WO200140294-A1.

XX 07-JUN-2001.

XX 27-NOV-2000; 2000WO-CN000494.

XX 29-NOV-1999; 99CN-00124148.

XX (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2001-397948/42.

XX N-PSDB; AAH45131.

XX Human thrombospondin-30 and polynucleotide is useful in diagnosis and  
 PT treatment of, e.g., malignant tumor, hemopathy, HIV infection,  
 PT immunological diseases and various inflammatory diseases.

XX Claim 1; Page 27-28; 33pp; Chinese.

CC The present sequence is the protein sequence for human thrombospondin-30.  
 CC Thrombospondin-30 protein and coding sequence are useful in the diagnosis  
 CC and treatment of malignant tumour, haemopathy, HIV infection,  
 CC immunological diseases and various inflammation diseases. In addition  
 CC thrombospondin-30 protein may be used for screening mimics, agonists,  
 CC antagonists or inhibitors, or for use in peptide fingerprinting  
 CC identification. The thrombospondin-30 coding sequence may be used as  
 CC primers for nucleic acid amplification reaction or as probes for  
 CC hybridisation reaction, or in producing gene chips or microarrays  
 XX  
 SQ Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 4; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKPRL 62  
 DB 2 HLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKPRL 61

QY 63 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 122  
 DB 62 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 121

QY 123 LGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTTETRVREI 182  
 DB 122 LGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTTETRVREI 181

QY 183 HPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNKGESKEAIPDSKSL 242  
 DB 182 HPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNKGESKEAIPDSKSL 241

QY 243 KEIPEQRNKKQKKRKVDKQKSVSVSTVH 273

DB 242 KEIPEQRNKKQKKRKVDKQKSVSVSTVH 272

## RESULT 4

AAM78328  
 ID AAM78328 standard; protein; 272 AA.

AC AAM78328;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 990.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK51461.  
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX Claim 20; Page 3214-3215; 6221pp; English.  
 PS The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAK78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAK80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX Sequence 272 AA;  
 SQ  
 Query Match 99.3%; Score 1505; DB 4; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HURLISWLFILNFMFYIGSQNASRRRRMHPNVSQCGGCATCSDYNGCLSKPRL 62  
 DB 2 HURLISWLFILNFMFYIGSQNASRRRRMHPNVSQCGGCATCSDYNGCLSKPRL 61  
 QY 63 PFALERIGMKQIGVCLSSPCSGYGYTRYPDINKTKCRADCTCFNKNFCTCKSGFYHL 122  
 DB 62 PFALERIGMKQIGVCLSSPCSGYGYTRYPDINKTKCRADCTCFNKNFCTCKSGFYHL 121  
 QY 123 LQKCLDNCPEGLAANNHTWCYSIVHCEVSEWNPSPCTKCKTGFGRGTFRVREIIQ 182  
 DB 122 LQKCLDNCPEGLAANNHTWCYSIVHCEVSEWNPSPCTKCKTGFGRGTFRVREIIQ 181  
 QY 183 HPSAKGNLCPTNTRKCTVQRKKQKGRGKGRKRRKKNKESKEAI PDSKLSRSS 242  
 DB 182 HPSAKGNLCPTNTRKCTVQRKKQKGRGKGRKRRKKNKESKEAI PDSKLSRSS 241  
 QY 243 KEIPQRENKQKQKRVQDKQKSVSVTVH 273  
 DB 242 KEIPQRENKQKQKRVQDKQKSVSVTVH 272  
 RESULT 5  
 AAE13168  
 ID AAE13168 standard; protein; 272 AA.  
 AC AAE13168;  
 XX 28-JAN-2002 (first entry)  
 XX Human stem cell growth factor-like protein #4.  
 DE Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;  
 KW neuroprotective; vulnary; cyrostatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;

KW supporting factor for the proliferation of stem cell.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /label= Signal\_peptide  
 FT Protein 22..272 /note= "Human mature stem cell growth factor-like  
 FT protein"  
 XX WO200177169-A2.  
 XX 18-OCT-2001.  
 PD 05-APR-2001; 2001WO-US011208.  
 XX 05-APR-2000; 2000US-00543774.  
 XX 28-JUN-2000; 2000US-0215733P.  
 PR 09-JAN-2001; 2001US-00757562.  
 PR 05-FEB-2001; 2001US-0266614P.  
 XX (HYSE-) HYSEQ INC.  
 PA (KIRI ) KIRIN BEER KK.  
 XX Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;  
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
 XX WPI; 2001-657166/75.  
 DR N-PSDB; AAD21728.  
 XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 FT disease, cancer, rheumatoid arthritis, osteoporosis.  
 XX Claim 28; Page 226-227; 232pp; English.  
 XX The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumours. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 CC inflammation. Sequences of the invention are also useful in gene therapy.  
 CC The present sequence is stem cell growth factor-like protein from human  
 XX Sequence 272 AA;  
 SQ  
 Query Match 99.3%; Score 1505; DB 4; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HURLISWLFILNFMFYIGSQNASRRRRMHPNVSQCGGCATCSDYNGCLSKPRL 62  
 DB 2 HURLISWLFILNFMFYIGSQNASRRRRMHPNVSQCGGCATCSDYNGCLSKPRL 61  
 QY 63 PFALERIGMKQIGVCLSSPCSGYGYTRYPDINKTKCRADCTCFNKNFCTCKSGFYHL 122



Db 62 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCYKCKSGFYH 121  
 QY 123 LGKCLDNCPEGLBANNHTECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTRVREIIQ 182  
 Db 122 LGKCLDNCPEGLBANNHTECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTRVREIIQ 181  
 QY 183 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 242  
 Db 182 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 241  
 QY 243 KEIPEORENKQKKRKVKQDKQKSVSVSTVH 273  
 Db 242 KEIPEORENKQKKRKVKQDKQKSVSVSTVH 272

## RESULT 6

AAE13150  
ID AAE13150 standard; protein; 272 AA.

AC AAE13150;

XX 28-JAN-2002 (first entry)

XX Human stem cell growth factor-like protein #2.

KW Human; stem cell growth factor-like protein; antiinflammatory; neutropic;  
 KW neuroprotective; vulnerary; cytostatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 KW supporting factor for the proliferation of stem cell.

XX Homo sapiens.

XX Key Location/Qualifiers

PH Peptide 1..21

FT /label= signal\_peptide

FT Protein 22..272

FT /note= "Human mature stem cell growth factor-like protein"

XX WO200177169-A2.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX 05-APR-2000; 2000US-00543774.

XX 28-JUN-2000; 2000US-0215733P.

XX 09-JAN-2001; 2000US-00757562.

XX 05-FEB-2001; 2001US-0266614P.

XX (HYSE-) HYSEO INC.

XX (KIRI) KIRIN BEER KK.

XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;

XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

XX WPI; 2001-657166/75.

XX N-PSDB; AAD21724.

XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.

PS Claim 28; Page 211-212; 232pp; English.

XX The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumours. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 CC inflammation. Sequences of the invention are also useful in gene therapy.  
 CC The present sequence is stem cell growth factor-like protein from human

XX Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. No. 2.8e-110;

Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HLELISWLPILFMFMEYIGSNASRGRRORRMRHNPVNSQGGCGCATCDYNGCLSKPRL 62

Db 2 HLELISWLPILFMFMEYIGSNASRGRRORRMRHNPVNSQGGCGCATCDYNGCLSKPRL 61

QY 63 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCYKCKSGFYH 122

Db 62 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCYKCKSGFYH 121

QY 123 LGKCLDNCPEGLBANNHTECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTRVREIIQ 182

Db 122 LGKCLDNCPEGLBANNHTECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTRVREIIQ 181

QY 183 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 242

Db 182 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 241

QY 243 KEIPEORENKQKKRKVKQDKQKSVSVSTVH 273

Db 242 KEIPEORENKQKKRKVKQDKQKSVSVSTVH 272

## RESULT 7

ABR62112

ID ABR62112 standard; protein; 272 AA.

XX ABR62112;

XX 18-AUG-2003 (first entry)

XX Human clone 1 thrombospondin protein #23.

KW Human; secreted; stem cell growth factor; cytostatic; haemostatic;  
 KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;  
 KW degenerative disease; Alzheimer's disease; food supplement;  
 KW immunological disorder; thrombospondin.

XX Homo sapiens.

XX WO2003029405-A2.

PD 10-APR-2003.  
XX 30-AUG-2002; 2002WO-US027746.  
XX 30-AUG-2001; 2001US-0316368P.  
PR 10-DEC-2001; 2001US-0339739P.  
PR 19-APR-2002; 2002US-00125852.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT;  
XX WPI; 2003-381616/36.  
XX New stem cell growth factor-like polypeptides and polynucleotides, useful  
PT for treating e.g. leukemia, hemophilia and degenerative diseases like  
PT Alzheimer's disease, and for inducing immune response.  
XX Disclosure; Fig 1; 151pp; English.  
XX The invention relates to new stem cell growth factor-like polypeptides  
CC and polynucleotides. The stem cell growth factor-like polypeptides and  
CC polynucleotides are useful for inducing differentiation of embryonic and  
CC adult stem cells to give rise to different cell types, for treating e.g.  
CC leukemia, haemophilia and degenerative diseases like Alzheimer's  
CC disease. They are also useful for generating new tissues and organs that  
CC may aid patients in need of transplanted tissues. The polynucleotides are  
CC useful as hybridisation probes, oligomers or primers for PCR, for  
CC chromosome and gene mapping, in recombinantly producing protein, in  
CC generating antisense DNA or RNA, in diagnostics as expressed sequence  
CC tags for identifying expressed genes, and for inducing immune response.  
CC The polypeptides are useful for generating antibodies that specifically  
CC bind the polypeptide, as molecular weight markers, and as a food  
CC supplement (e.g. protein or amino acid supplement, and as a carbon,  
CC nitrogen or carbohydrate source). Compositions comprising the  
CC polypeptides or polynucleotides are useful for the diagnosis, treatment  
CC or prevention of cancers, and other immunological disorders. The current  
CC sequence represents a human clone 1 thrombospondin protein  
XX  
XX Sequence 272 AA;  
XX  
XX Query Match 99.3%; Score 1505; DB 6; Length 272;  
XX Best Local Similarity 100.0%; Pred. No. 2.8e-110;  
XX Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 HURLISWLPILNFMFYIGSQNASRGRQRMRHPNVSGQGCGCATCSYNGCLCKPRL 62  
Db 2 HURLISWLPILNFMFYIGSQNASRGRQRMRHPNVSGQGCGCATCSYNGCLCKPRL 61  
QY 63 PFALERIGMKOIGVCLSSCPGYYGTRYPDINKTKCKADCDTCFNKNFTCKCKSGFYLLH 122  
Db 62 PFALERIGMKOIGVCLSSCPGYYGTRYPDINKTKCKADCDTCFNKNFTCKCKSGFYLLH 121  
QY 123 LGKCLDNCPEGLEANNHTMECVSIVHCVSEWNPWSPTCKGKTCGFRGTETRVREIIQ 182  
Db 122 LGKCLDNCPEGLEANNHTMECVSIVHCVSEWNPWSPTCKGKTCGFRGTETRVREIIQ 181  
QY 183 HPSAKGNLCPTNETRKTQVRKKCKGKGRKKGRKKRKKPNKNGSKSAIPDPSKLSLS 242  
Db 182 HPSAKGNLCPTNETRKTQVRKKCKGKGRKKGRKKRKKPNKNGSKSAIPDPSKLSLS 241  
QY 243 KEIPEORENKQKKKRVDDKQKSVSVTVH 273  
Db 242 KEIPEORENKQKKKRVDDKQKSVSVTVH 272  
XX  
XX RESULT 8  
XX ABO44432  
XX ID ABO44432 standard; protein; 272 AA.  
XX AC ABO44432;  
XX DT 30-SEP-2003 (first entry)

XX Human stem cell growth factor-like protein, SCR 1 #4.  
XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
KW immunostimulant; vulnery; haematopoietic stem cell; gene therapy;  
KW supporting factor for proliferation of stem cells; wound healing;  
KW haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;  
KW bone marrow transplantation; cord blood transplantation;  
KW chronic granulomatous disease; duplicated immunodeficiency syndrome;  
KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;  
KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;  
KW epithelial cell growth; ovarian follicle development; nerve cell growth;  
KW cartilage remodeling; bone growth; immunosuppression; human.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Peptide 1..21 /label= Signal\_peptide  
XX Protein 22..272 /note= "Mature stem cell growth factor-like protein. This  
XX protein is specifically claimed in claim 9"  
XX US2003044792-A1.  
XX 06-MAR-2003.  
XX 28-JUN-2001; 2001US-00894912.  
XX 28-JUN-2000; 2000US-0215733P.  
PR 05-FEB-2001; 2001US-0266614P.  
PR 05-APR-2001; 2001US-0282397P.  
XX (TANG/) TANG Y T.  
PA (LABA/) LABAT I.  
PA (DRMA/) DRMANAC R T.  
PA (MIZE/) MIZE N.  
PA (NISH/) NISHIKAWA M.  
PA (CHAO/) CHAO C.  
XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;  
XX WPI; 2003-625403/59.  
XX N-PSDB; ACH04328.  
XX Novel isolated polypeptide having stem cell growth factor activity,  
PT useful for promoting wound healing, and as a medicine to proliferate or  
PT support human hematopoietic stem cells or human hematopoietic progenitor  
PT cells.  
XX Claim 23; Page 82; 96pp; English.  
XX The invention relates to an isolated stem cell growth factor-like  
CC polypeptide (referred as supporting factor for proliferation of stem  
CC cells (SCR-1)) from mouse or human, or its mature protein portion, or  
CC fragment, analogue, variant or derivative, that retains stem cell growth  
CC factor activity. Also included are an isolated polynucleotide encoding  
CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or  
CC derivative, that retains stem cell growth factor activity, or the  
CC complement of the polynucleotide), an (expression) vector comprising the  
CC SCR-1 polynucleotide, a host cell genetically engineered to contain the  
CC SCR-1 polynucleotide in operative association with a regulatory sequence  
CC that controls expression of the polynucleotide in the host cell,  
CC preparation of the SCR-1 polypeptide, a polypeptide which is an  
CC expression product of the SCR-1 polynucleotide (the polypeptide having an  
CC activity to support proliferation or survival of haematopoietic stem cell  
CC or haematopoietic progenitor cell, with a proviso that C-terminal aa  
CC sequence does not comprise the aa sequence appearing as ABO44433), an  
CC isolated SCR-1 polypeptide with stem cell growth factor activity and  
CC lacking any 10 consecutive aa from ABO44430, an isolated polypeptide  
CC with stem cell growth factor activity having at least an aa sequence  
CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1  
CC polypeptide to maintain survival of or promote proliferation of a stem





Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; human.

OS Homo sapiens.

XX US2003044792-A1.

XX 06-MAR-2003.

XX 28-JUN-2001; 2001US-00894912.

XX 28-JUN-2000; 2000US-0215733P.

PR 05-FEB-2001; 2001US-0266614P.

PR 05-APR-2001; 2001US-0282397P.

XX (TANG/) TANG Y T.

PA (LABA/) LABAT I.

PA (DRMA/) DRMANAC R T.

PA (MIZE/) MIZE N.

PA (NISH/) NISHIKAWA M.

PA (CHAO/) CHAO C.

XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

XX WPI; 2003-625403/59.

XX Novel isolated polypeptide having stem cell growth factor activity.

XX useful for promoting wound healing, and as a medicine to proliferate or

XX support human hematopoietic stem cells or human hematopoietic progenitor

XX cells.

XX Disclosure; Fig 3; 96pp; English.

XX The invention relates to an isolated stem cell growth factor-like

XX polypeptide (referred as supporting factor for proliferation of stem

XX cells (SCR-1)) from mouse or human, or its mature protein portion, or

XX fragment, analogue, variant or derivative, that retains stem cell growth

XX factor activity. Also included are an isolated polynucleotide encoding

XX SCR-1 (or its mature protein portion, or fragment, analogue, variant or

XX derivative, that retains stem cell growth factor activity, or the

XX complement of the polynucleotide), an (expression) vector comprising the

XX SCR-1 polynucleotide, a host cell genetically engineered to contain the

XX SCR-1 polynucleotide in operative association with a regulatory sequence

XX that controls expression of the polynucleotide in the host cell,

XX preparation of the SCR-1 polypeptide, a polypeptide which is an

XX expression product of the SCR-1 polynucleotide (the polypeptide having an

XX activity to support proliferation or survival of haematopoietic stem cell

XX or haematopoietic progenitor cell, with a proviso that C-terminal aa

XX sequence does not comprise the aa sequence appearing as ABO44433), an

XX isolated SCR-1 polypeptide with stem cell growth factor activity and

XX lacking any 10 consecutive aas from ABO44430, an isolated polypeptide

CC healing. The human haematopoietic stem cell or human haematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of haematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia  
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's  
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth, and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a protein  
 CC homologous to Human SCR-1 protein  
 XX

SQ Sequence 265 AA;

Query Match 97.1%; Score 1472; DB 6; Length 265;

Best Local Similarity 100.0%; Pred. No. 1.1e-107; Indels 0; Gaps 0;

Matches 264; Conservative 0; Mismatches 0;

Qy 3 HLRLLSWLFIILNFMVYIGSQNASRRQRHHPNVSGCGGCATCSDYNGCLSKPRL 62

Db 2 HLRLLSWLFIILNFMVYIGSQNASRRQRHHPNVSGCGGCATCSDYNGCLSKPRL 61

Qy 63 PFALERIGMKOIGVCLSSCPGSGYTGTRYPDINKTKKADCDTCFNKQFCTKCKSGFYIH 122

Db 62 PFALERIGMKOIGVCLSSCPGSGYTGTRYPDINKTKKADCDTCFNKQFCTKCKSGFYIH 121

Qy 123 LGKCLDNCPEGLRANNHTMECVSIHVCEVSWNPWSPCTKKGKTCGPKRGTRVREIIQ 182

Db 122 LGKCLDNCPEGLRANNHTMECVSIHVCEVSWNPWSPCTKKGKTCGPKRGTRVREIIQ 181

Qy 183 HPSAKGNLCPTNETRKTCTVQRKCKGGRGKGRERKRKPKNGESKEAIPDSKSLESS 242

Db 182 HPSAKGNLCPTNETRKTCTVQRKCKGGRGKGRERKRKPKNGESKEAIPDSKSLESS 241

Qy 243 KEIPEQRNKKQKKRKKVQDKQKS 266

Db 242 KEIPEQRNKKQKKRKKVQDKQKS 265

RESULT 12

AAW85607

ID AAW85607 standard; protein; 292 AA.

XX AAW85607;

XX 02-MAR-1999 (first entry)

XX Secreted protein clone da228\_6.

XX Clone; secreted protein; protein factor; cytokine; lymphokine;

XX interferon; colony stimulating factor; CSF; interleukin; cloning;

XX tumour invasion; tumour suppression; immune boosting.

XX Homo sapiens.

XX WO9849302-A1.

XX 05-NOV-1998.

XX 24-APR-1998; 98WO-US008336.

XX 25-APR-1997; 97US-00845296.

XX 23-APR-1998; 98US-00065125.

XX (GENY ) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;

XX Spaulding V, Agostino MJ;

XX WPI; 1999-024059/02.

DR N-PSDB; AAV83133.  
 XX New polynucleotides encoding secreted human proteins - are derived from  
 PT human foetal brain, adult brain, adult blood or placenta cDNA libraries,  
 PT useful, e.g. as potential immunomodulators.  
 XX Claim 8; Page 63-64; 104pp; English.  
 XX The nucleotide sequence (NS) of the full-length protein-coding sequence  
 CC of clones c1254 (AAV83132), da2286 (AAV83133), du4105 (AAV83134), eh801  
 CC (AAV83135), er3691 (AAV83136), fh1235 (AAV83137), fm601 (AAV83138) or  
 CC fr4732 (AAV83139). (all clones are deposited as ATCC 98415) and the  
 CC proteins they encode are predicted to have biological activities which  
 CC would make them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals for example, tumour suppression/invasion  
 CC activity, immune system boosting activity. The polynucleotides are also  
 CC believed to be useful for gene therapy  
 XX Sequence 292 AA;  
 XX  
 XX Query Match 97.1%; Score 1472; DB 2; Length 292;  
 XX Best Local Similarity 100.0%; Pred. No. 1.2e-107;  
 XX Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HLELISWLFILFMFYIGSONASRGRRRRRHPNVSQGGCGATCSDYNGCLSCPKRL 62  
 DB 2 HLELISWLFILFMFYIGSONASRGRRRRRHPNVSQGGCGATCSDYNGCLSCPKRL 61  
 QY 63 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCDTCFNNKFTCKSGPYLH 122  
 DB 62 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCDTCFNNKFTCKSGPYLH 121  
 QY 123 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKGGTKCGFRGTETVRRIIQ 182  
 DB 122 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKGGTKCGFRGTETVRRIIQ 181  
 QY 183 HPSAKGNLCPTTNETRKTCTVQRKCKGGRKGRKRRKPKNGSKRAIPDSKLESS 242  
 DB 182 HPSAKGNLCPTTNETRKTCTVQRKCKGGRKGRKRRKPKNGSKRAIPDSKLESS 241  
 QY 243 KEIPEORENKQOQKGRKVDQKQS 266  
 DB 242 KEIPEORENKQOQKGRKVDQKQS 265  
 RESULT 13  
 ID AAE13170  
 XX AAE13170 standard; protein; 292 AA.  
 XX AC AAE13170;  
 XX DT 28-JAN-2002 (first entry)  
 XX DE Human SCR-1 related protein.  
 XX KW Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;  
 KW neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disease;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 KW supporting factor for the proliferation of stem cell.  
 XX Unidentified.  
 XX OS WO200177169-A2.  
 XX PN  
 XX

PD 18-OCT-2001.  
 XX 05-APR-2001; 2001WO-US011208.  
 XX 05-APR-2000; 2000US-00543774.  
 PR 28-JUN-2000; 2000US-0215733P.  
 PR 09-JAN-2001; 2001US-00757562.  
 PR 05-FEB-2001; 2001US-0266614P.  
 XX (HYSB-) HYSQ INC.  
 PA (KIRI ) KIRIN BEER KK.  
 XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;  
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
 XX WPI; 2001-657166/75.  
 DR N-PSDB; AAD21740.  
 XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.  
 XX Claim 27; Page 231-232; 232pp; English.  
 PS The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumours. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 CC inflammation. Sequences of the invention are also useful in gene therapy.  
 CC The present sequence is human SCR-1 related protein  
 XX Sequence 292 AA;  
 XX  
 XX Query Match 97.1%; Score 1472; DB 4; Length 292;  
 XX Best Local Similarity 100.0%; Pred. No. 1.2e-107;  
 XX Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HLELISWLFILFMFYIGSONASRGRRRRRHPNVSQGGCGATCSDYNGCLSCPKRL 62  
 DB 2 HLELISWLFILFMFYIGSONASRGRRRRRHPNVSQGGCGATCSDYNGCLSCPKRL 61  
 QY 63 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCDTCFNNKFTCKSGPYLH 122  
 DB 62 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCDTCFNNKFTCKSGPYLH 121  
 QY 123 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKGGTKCGFRGTETVRRIIQ 182  
 DB 122 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKGGTKCGFRGTETVRRIIQ 181  
 QY 183 HPSAKGNLCPTTNETRKTCTVQRKCKGGRKGRKRRKPKNGSKRAIPDSKLESS 242  
 DB 182 HPSAKGNLCPTTNETRKTCTVQRKCKGGRKGRKRRKPKNGSKRAIPDSKLESS 241  
 QY 243 KEIPEORENKQOQKGRKVDQKQS 266  
 XX



Db 242 KEIPEQRENKQQKKRVQDKQKS 265

RESULT 14  
ABP61846  
ID ABP61846 standard; protein; 292 AA.  
AC ABP61846;  
XX  
DT 04-OCT-2002 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 200.  
XX  
KW Human, cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;  
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;  
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;  
KW antitumor; fungicide; antidiabetic; antiaesthetic; antiallergic;  
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;  
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;  
KW stem cell; growth factor; nervous system disease; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW osteoporosis; severe combined immunodeficiency; SCID; infection;  
KW multiple sclerosis; rheumatoid arthritis; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN US2002065394-A1.  
XX  
XX 30-MAY-2002.  
XX  
XX 22-DEC-2000; 2000US-00745763.  
XX  
XX 18-MAR-1998; 98US-00040963.  
XX  
XX (JACO/) JACOBS K.  
XX (MCCO/) MCCOY J M.  
XX (LAVA/) LAVALLIE E R.  
XX (COLL/) COLLINS-RACIE L A.  
XX (EVAN/) EVANS C.  
XX (MERB/) MERBERG D.  
XX (TREA/) TREACY M.  
XX (SPAU/) SPAULDING V.  
XX  
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Spaulding V;  
XX  
XX WPI; 2002-582343/62.  
XX N-PSDB; ABQ92060.  
XX  
XX Novel secreted or transmembrane protein and polynucleotide encoding the  
XX protein, useful for diagnosis and treatment of neurological disorders,  
XX cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.  
XX  
XX Claim 207; Page 203-204; 284pp; English.  
XX  
XX The invention relates to human secreted or transmembrane protein (I),  
XX their fragments and is encoded by specific complementary deoxyribonucleic  
XX acid (cDNA) inserts (II), where the protein is substantially free from  
XX other mammalian proteins. (I) are useful for preventing, treating or  
XX ameliorating a medical condition, especially immunological treatment or  
XX prevention of tumours. (I) exhibits activity relating to angiogenesis,  
XX cytokine, cell proliferation, cell differentiation, antiinflammatory,  
XX stem cell growth factor activity and activin or inhibin-related  
XX activities. (I) can be used to manipulate stem cells in culture to give  
XX rise to neuroepithelial cells that can be used to augment or replace  
XX cells damaged by illnesses, autoimmune disease, accidental damage or  
XX genetic disorders. (I) induces the proliferation of neural cells and  
XX regeneration of nerve and brain tissue and is useful for the treatment of  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis. (I) is involved in chemotactic or chemokinetic  
XX activity, regulation of haematopoiesis and is useful for treating myeloid  
XX or lymphoid cell disorders, platelet disorders such as thrombocytopaenia

CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers, or  
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or  
CC periodontal disease. (I) is also useful for gut protection or  
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury  
CC in various tissues, various immune deficiencies and disorders including  
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,  
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,  
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,  
CC such as asthma or other respiratory problems. (II) is useful to express  
CC recombinant protein, as markers for tissues in which the corresponding  
CC protein is preferentially expressed and in gene therapy. The present  
CC sequence is that of a polypeptide of the invention  
XX  
SQ Sequence 292 AA;  
Query Match 97.1%; Score 1472; DB 5; Length 292;  
Best Local Similarity 100.0%; Pred. No. 1.2e-107;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 HLRILSWLPIILNPMVEYIGSQNASRGRORRHPNVSGQCGGCATCSDYNGCLSCXPRL 62  
Db 2 HLRILSWLPIILNPMVEYIGSQNASRGRORRHPNVSGQCGGCATCSDYNGCLSCXPRL 61  
Qy 63 PFALERIGMKQIGVCLSSCPSSGVYGYTRYPDINKCTKCKADCCTCFNKNFCTKCKSGFYLN 122  
Db 62 PFALERIGMKQIGVCLSSCPSSGVYGYTRYPDINKCTKCKADCCTCFNKNFCTKCKSGFYLN 121  
Qy 123 LGKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPWSPCTKXGKTGCFXRGTTETRVREIIQ 182  
Db 122 LGKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPWSPCTKXGKTGCFXRGTTETRVREIIQ 181  
Qy 183 HPSAKGNLCPPTNTRCTVQKKGKGRGKGRKRRKPKNGESKEAIPDSKSLSS 242  
Db 182 HPSAKGNLCPPTNTRCTVQKKGKGRGKGRKRRKPKNGESKEAIPDSKSLSS 241  
Qy 243 KEIPEQRENKQQKKRVQDKQKS 266  
Db 242 KEIPEQRENKQQKKRVQDKQKS 265  
RESULT 15  
ABR62114  
ID ABR62114 standard; protein; 292 AA.  
XX  
AC ABR62114;  
XX  
DT 18-AUG-2003 (first entry)  
XX  
DE Human secreted protein clone da\_288\_6 #25.  
XX  
KW Human; secreted; stem cell growth factor; cytostatic; haemostatic;  
KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;  
KW degenerative disease; Alzheimer's disease; food supplement;  
KW immunological disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029405-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 30-AUG-2002; 2002WO-US027746.  
XX  
PR 30-AUG-2001; 2001US-0316368P.  
PR 10-DEC-2001; 2001US-0339739P.  
PR 19-APR-2002; 2002US-00125852.  
XX  
PA (HYSB-) HYSBQ INC.  
XX  
PI Tang YT;  
XX  
DR WPI; 2003-381616/36.



XX New stem cell growth factor-like polypeptides and polynucleotides, useful  
PT for treating e.g. leukemia, hemophilia and degenerative diseases like  
PT Alzheimer's disease, and for inducing immune response.  
XX  
XX  
PS Disclosure; Fig 2; 151pp; English.  
PS  
XX The invention relates to new stem cell growth factor-like polypeptides  
XX and polynucleotides. The stem cell growth factor-like polypeptides and  
CC polynucleotides are useful for inducing differentiation of embryonic and  
CC adult stem cells to give rise to different cell types, for treating e.g.  
CC leukaemia, haemophilia and degenerative diseases like Alzheimer's  
CC disease. They are also useful for generating new tissues and organs that  
CC may aid patients in need of transplanted tissues. The polynucleotides are  
CC useful as hybridisation probes, oligomers or primers for PCR, for  
CC chromosome and gene mapping, in recombinantly producing protein, in  
CC generating antisense DNA or RNA, in diagnostics as expressed sequence  
CC tags for identifying expressed genes, and for inducing immune response.  
CC The polypeptides are useful for generating antibodies that specifically  
CC bind the polypeptide, as molecular weight markers, and as a food  
CC supplement (e.g. protein or amino acid supplement, and as a carbon,  
CC nitrogen or carbohydrate source). Compositions comprising the  
CC polypeptides or polynucleotides are useful for the diagnosis, treatment  
CC or prevention of cancers, and other immunological disorders. The current  
CC sequence represents a human secreted protein clone da\_288\_6  
XX  
XX Sequence 292 AA;

Query Match 97.1%; Score 1472; DB 6; Length 292;  
Best Local Similarity 100.0%; Pred. No. 1.2e-107;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HLELIISWLFIIINMEYIGSNASRGRRORRHPNVSQGGCGATCSYNGCLCKPRL 62  
DB |||||  
QY 2 HLELIISWLFIIINMEYIGSNASRGRRORRHPNVSQGGCGATCSYNGCLCKPRL 61  
DB |||||  
QY 63 PFALERIGMKQIGVCLSSCPSPGYGYTRYPDINKCTCKADCTCFNKNFCTCKSGFYLIH 122  
DB |||||  
QY 62 PFALERIGMKQIGVCLSSCPSPGYGYTRYPDINKCTCKADCTCFNKNFCTCKSGFYLIH 121  
DB |||||  
QY 123 LGKCLDNCPEGLBANNHMECVSIHCEVSENNPWSPTKKGKTCGFGTETRVREIIQ 182  
DB |||||  
QY 183 HPSAKGNLCPTNETRKCTVQRKCKQGRGKGRERKRNKNGESKEAIPDSKSLSS 242  
DB |||||  
QY 182 HPSAKGNLCPTNETRKCTVQRKCKQGRGKGRERKRNKNGESKEAIPDSKSLSS 241  
DB |||||  
QY 243 KEIPEORENKQOKRKRKVDQKKS 266  
DB |||||  
QY 242 KEIPEORENKQOKRKRKVDQKKS 265  
DB |||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:57:05 ; Search time 14.2464 Seconds  
(without alignments)  
989.298 Million cell updates/sec

Title: US-09-894-912A-13  
Perfect score: 1516  
Sequence: 1 MGHRLISMLFIILNFEYI.....QQKKRKVDKQKSVSVTVH 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	173.5	11.4	969	2	US-08-284-941-2
2	173.5	11.4	969	2	US-08-447-642-2
3	173.5	11.4	969	3	US-09-236-503-2
4	173.5	11.4	969	5	PCT-US93-02147A-2
5	162	10.7	799	2	US-08-525-940-23
6	162	10.7	799	2	US-08-976-838-23
7	162	10.7	881	2	US-08-525-940-21
8	162	10.7	881	2	US-08-976-838-21
9	162	10.7	915	2	US-08-525-940-18
10	162	10.7	915	2	US-08-976-838-18
11	162	10.7	915	4	US-09-214-555B-2
12	162	10.7	915	4	US-09-214-555B-7
13	161	10.6	288	1	US-08-368-852-15
14	158.5	10.5	379	4	US-09-907-794A-4
15	158.5	10.5	379	4	US-09-905-125A-4
16	158.5	10.5	379	4	US-09-302-775A-4
17	157.5	10.4	288	2	US-08-525-940-15
18	157.5	10.4	288	2	US-08-976-838-15
19	150.5	9.9	380	4	US-09-205-258-441
20	149	9.8	568	1	US-07-862-021B-14
21	149	9.8	568	5	PCT-US93-03164-14
22	148.5	9.8	2523	1	US-08-185-432-18
23	148.5	9.8	2523	4	US-08-899-232-3
24	147.5	9.7	802	1	US-07-862-021B-12
25	147.5	9.7	802	1	US-08-313-288B-12
26	147.5	9.7	802	5	PCT-US93-03164-12
27	146.5	9.7	807	4	US-09-132-769-1

28	146.5	9.7	807	4	US-09-640-173-186	Sequence 186, Appl
29	146.5	9.7	807	4	US-09-713-550-186	Sequence 186, Appl
30	146	9.6	1068	1	US-08-537-210A-2	Sequence 2, Appli
31	146	9.6	1068	3	US-09-113-825-2	Sequence 2, Appli
32	146	9.6	2556	1	US-08-185-432-17	Sequence 17, Appli
33	146	9.6	2556	4	US-08-899-232-2	Sequence 2, Appli
34	144	9.5	807	4	US-09-132-769-3	Sequence 3, Appli
35	143.5	9.5	807	1	US-07-862-021B-10	Sequence 10, Appl
36	143.5	9.5	807	1	US-08-313-288B-10	Sequence 10, Appl
37	143.5	9.5	807	4	US-09-132-769-5	Sequence 5, Appli
38	143.5	9.5	807	5	PCT-US93-03164-10	Sequence 103, Appl
39	139	9.2	366	3	US-08-857-076-103	Sequence 5, Appli
40	139	9.2	486	3	US-08-746-559A-5	Sequence 4, Appli
41	139	9.2	516	3	US-08-746-559A-4	Sequence 2, Appli
42	139	9.2	1367	2	US-08-249-887C-2	Sequence 2, Appli
43	139	9.2	1367	2	US-08-625-819-2	Sequence 2, Appli
44	139	9.2	1367	3	US-08-746-559A-2	Sequence 2, Appli
45	139	9.2	1367	4	US-08-864-641B-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-08-284-941-2  
; Sequence 2, Application US/08284941  
; Patent No. 5863756  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; APPLICANT: KIEFER, MICHAEL C  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND  
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/284,941  
; FILING DATE: 2 August 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEELLY PH.D., RICHARD L.  
; REGISTRATION NUMBER: 30092  
; REFERENCE/DOCKET NUMBER: CHIR-009/01US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 843-5070  
; TELEFAX: (415) 857-0683  
; TELEX: 380816 COOLY FA  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 969 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-284-941-2

Query Match 11.4%; Score 173.5; DB 2; Length 969;  
Best Local Similarity 28.8%; Pred. No. 1.2e-06;  
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;  
QY 19 YIGSQNASRGRQRMRHPNVSVQCQGGCATCSD--YNGCLSCKPRLPFALERIGMKQIGV 76  
Db 737 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGPF-----HHQSMNT 779



```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-02147A-2

Query Match      11.4%; Score 173.5; DB 5; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.2e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY   19 YIGSQNASRRQRMRHPNVSQCQGSCATCSD--YNGCLSKPRLPFFALERIMKQIGV 76
DB   737 YFDYIARRCR-----CHKGCTCSSRAATQLCSR-RGFY-----HHQMWT 779

QY   77 CLSSCPGSGYGYTRYPDINKTKCKADCDTCFNK-NFTCKSGFYHLGLKCLDNCPGLE 135
DB   780 CVTLCPAGPYADE-SQKNCLKHPCPKCVDPKCTVKCEGSFLARGSCIPDCPEGVY 837

QY   136 ANHTMECVSIHVCEVSEWNPSPCTKKGTGG 168
DB   838 FDSLRIRCGECH-----TOG 853

RESULT 5
US-08-525-940-23
; Sequence 23, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; TITLE OF INVENTION: CD4+ T-LYMPHO CYTE PROTEASES AND GENES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid

; Query Match      10.7%; Score 162; DB 2; Length 799;
; Best Local Similarity 22.3%; Pred. No. 9.4e-06;
; Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-23

Query Match      10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.4e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY   21 GSONASRGRRMRHPNVSQCQGSCATCSDYNG--CLSKPRLPFPALBRIGHMKQIGVCL 78
DB   563 GHYHADK-KRCRKCAPN-----CESCFGSHGDQCMASKCYGYL-----NEETNSCV 607

QY   79 SCPSG-YGYTRYPDINKTKCKADCDTCFNKFCFKSGFYHLGLKCLDNCPGLEBAN 137
DB   608 THCPGSGYQDTKK---NLCKRCSENCKTCTEPHNCTECRDGLSLQSGRCSVCEDGRYFN 664

QY   138 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 163
DB   665 GDQPCHRFPCATCAGACDGICNTGEYFMEDGRCVQSCSISYYFDHSSENGYSCKKC 724

QY   164 GKTC-----GFRKRTG-----TRVREIIQHPSAKGNLCPTTNSTRKCTV 202
DB   725 DISCLTNGPGPFNKCTSCPSPGYLLDLGMQMGAI CKDATEBSWAEGGFCMLVKONLNC-- 782

QY   203 QKCKQK 209
DB   783 QRKVQQ 789

RESULT 6
US-08-976-838-23
; Sequence 23, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHO CYTE PROTEASE NUCLEIC ACID
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-23

; Query Match      10.7%; Score 162; DB 2; Length 799;
; Best Local Similarity 22.3%; Pred. No. 9.4e-06;
; Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

```



QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 163  
Db 747 GDCQCPCHRCATCAGAGAGCINCTEGYPMWDRGCVQSCSISYYFDHSSSENGYKSKKC 806  
QY 164 GRTG-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 202  
Db 807 DISCLTCNGPGFNCTSCPSGYLLDLGMCQGAICKDATESWAEAGGFCMLVKKNLNC-- 864  
QY 203 ORKKCQK 209  
Db 865 QKVLQO 871

## RESULT 9

US-08-525-940-18  
; Sequence 18, Application US/08525940  
; Patent No. 5866351  
; GENERAL INFORMATION:  
; APPLICANT: Franzusoff, Alex  
; APPLICANT: Miranda, Luis R.  
; APPLICANT: Wolf, Joseph R.  
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES  
; TITLE OF INVENTION: ENCODING SAID PROTEASES  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/525,940  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/366,852  
; FILING DATE: 01-JAN-1995  
; PRIOR APPLICATION DATA: US 08/088,322  
; APPLICATION NUMBER: US 08/088,322  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2848-11-C1  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-525-940-18

Query Match 10.7%; Score 162; DB 2; Length 915;  
Best Local Similarity 22.3%; Pred. No. 1.1e-05;  
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;  
QY 21 GSONASRGRRQRHMPNVSGCGGCATCSDYNG--CLSCPKPLFPALERIGMKQIGVCL 78  
Db 679 GHYHADK-KKCRKCAPN-----CBSCFGSHGDCQMSCKYGYFL-----NEETNSCV 723  
QY 79 SSCPSG-YGTRYPDINKTKCKADCDTCFNKFCFKSGFYHLGLKCLDNCPEGLEAN 137  
Db 724 THCPDGSYQDTTK--NLCKKSENCKTCTBFHNTCTCRDGLSLQSGRCVSCDGRYFN 780

QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 163  
Db 781 GDCQCPCHRCATCAGAGAGCINCTEGYPMWDRGCVQSCSISYYFDHSSSENGYKSKKC 840  
QY 164 GRTG-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 202  
Db 841 DISCLTCNGPGFNCTSCPSGYLLDLGMCQGAICKDATESWAEAGGFCMLVKKNLNC-- 898  
QY 203 ORKKCQK 209  
Db 899 QKVLQO 905

## RESULT 10

US-08-976-838-18  
; Sequence 18, Application US/08976838  
; Patent No. 5981259  
; GENERAL INFORMATION:  
; APPLICANT: Franzusoff, Alex  
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASE NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln St., Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,838  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2848-11-C2  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-976-838-18

Query Match 10.7%; Score 162; DB 2; Length 915;  
Best Local Similarity 22.3%; Pred. No. 1.1e-05;  
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;  
QY 21 GSONASRGRRQRHMPNVSGCGGCATCSDYNG--CLSCPKPLFPALERIGMKQIGVCL 78  
Db 679 GHYHADK-KKCRKCAPN-----CBSCFGSHGDCQMSCKYGYFL-----NEETNSCV 723

QY 79 SSCPSG-YGTRYPDINKTKCKADCDTCFNKFCFKSGFYHLGLKCLDNCPEGLEAN 137  
Db 724 THCPDGSYQDTTK--NLCKKSENCKTCTBFHNTCTCRDGLSLQSGRCVSCDGRYFN 780  
QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 163  
Db 781 GDCQCPCHRCATCAGAGAGCINCTEGYPMWDRGCVQSCSISYYFDHSSSENGYKSKKC 840  
QY 164 GRTG-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 202  
Db 841 DISCLTCNGPGFNCTSCPSGYLLDLGMCQGAICKDATESWAEAGGFCMLVKKNLNC-- 898

```
QY 203 QKCKQK 209
Db 899 QKVLQK 905

RESULT 11
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match 10.7%; Score 162; DB 4; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 76; Gaps 12;

QY 21 GSONASRGRRORRMPNVNSQGGCGCATCSDYNG--CLSKPRLFFALERIGMKQIGVCL 78
Db 679 GHYHADK-KRCRKCAPN-----CESCFGSHGDCMSCKYGYFL-----NEETNSCV 723

QY 79 SSCPSG-YGTRYPDINKTKKADCDTCFNKFNCTKCKSGFYHLGKCLDNCBPGLAN 137
Db 724 THCPDGSYQDTTK--NLCKRCSNCKYCTEFHCTECRDLGSLQGRCSVSCDGRYFN 780

QY 138 NHTMB-----CVS-----IVHCEVSEW-----NFWSPCTKK 163
Db 781 GDCQCPCHRFATCAGAGADGCTGEGYFMDGRCVQSCISYYPDHSSSENGYKCKKC 840

QY 164 GKTC-----GFKRGTE-----TRVREIIQHPSAKGNLCPTTNETRKCTV 202
Db 841 DISCLTCNGPGFKNCTSCPSGYLLDLGMCQGAICKDATEESWAGGFCMLVKKNLNC-- 898

QY 203 QKCKQK 209
Db 899 QKVLQK 905

RESULT 12
US-09-214-555B-7
; Sequence 7, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-7

Query Match 10.6%; Score 161; DB 1; Length 288;
Best Local Similarity 28.2%; Pred. No. 3.6e-06;
Matches 46; Conservative 25; Mismatches 52; Indels 40; Gaps 11;

QY 36 PNVSQ-GCGC-GCATCSDYNGCLSKPRLFFALERIGMKQIGVCLSSCPGYYGTGTPDI 93
```



Db 13 PECSEVCGDGPDPCHND---CL---HYKKL---NNTRICVSSCPGHY---HADK 58  
Qy 94 NKCTKACDCTPNK--NFTCKCKSGFYH--LGKCLDNCPEGLNANNHTMBCVSIHVC 149  
Db 59 KRCKKCAPNCESCFSGHQDQMSCKYGYFLMEETNSCVTHCPDGSQDITKKNLC-----112  
Qy 150 EVSEWNPWSPCTKKGKTC-GPKRGVTRVREIIHQHPSAKGNLC 191  
Db 113 -----RKSENKTCRTHFKCTECR-----DGLSLQGSRC 142

## RESULT 14

US-09-907-794A-4

; Sequence 4, Application US/09907794A  
; Patent No. 665468  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Deenoyers, Luc  
; APPLICANT: Baton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kijavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/907,794A  
; CURRENT FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095

; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 4  
; LENGTH: 379  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-907-794A-4

## Query Match

10.5%; Score 158.5; DB 4; Length 379;

Best Local Similarity 24.9%; Pred. No. 8.1e-06;

Matches 60; Conservative 72; Mismatches 72; Indels 83; Gaps 16;

Qy 42 CQGGCAT---CSDYNGC-----LSCKPRLFFALERIGMKQIGVCLSS---C 81  
Db 182 CPGCRNGGFCNERRICECPDPGFHGHPCFKALCTPRCHG-----GLCVTFPGCIC 232  
Qy 82 PSGYGYTRYPDINKCTKC-KADCD-TCFNKNFCTKCKSGFYHLGKCLDNCPEGLNANH 139  
Db 233 PPGFYG-----VNCCKANCSTTCFNGTC-----FY--PKCI--CPPGLEGB-- 271  
Qy 140 TMECVSIVHCEVSEWNPWSPCTKKGKTCGPKRGVTRVREIIHQHPSAKGNLCPP-----193  
Db 272 -----QCEISKCP--QPCRNGKCGIG---KSKCKSGYQGLDCKSPVCBPGCGAHG 318  
Qy 194 -TNETKCTVQRKKCKGKRGKGRERKRK-----KPNKGSKEAIPDSLSLESSKEIP 246  
Db 319 TCHEPNKC-----QCQGWGHRHCNRYEASLIHALRPAQLRQHTPSLKAERRDPP 373  
Qy 247 E 247  
Db 374 E 374

## RESULT 15

US-09-905-125A-4

; Sequence 4, Application US/09905125A

; Patent No. 6664376

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Deenoyers, Luc  
; APPLICANT: Baton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kijavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/905,125A

```
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 4
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-125A-4

Query Match      10.5%; Score 158.5; DB 4; Length 379;
Best Local Similarity 24.9%; Pred. No. 8.1e-06;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Qy      42  CQGGCAT---CSDYNGC-----LSCKPRLFPALERIGMKQIGVCLSS-----C 81
Db      182  CPGGCRNGGFCNERICECPDGFHGHPCBKALCTPCNNG-----GLCVTPGFCIC 232

Qy      82  PSGYTGTRYPDINKTKC-KAPCD-TCFNKFNCTKCKSGFYHLGKCLDNCFEGLEANNH 139
Db      233  PRGFYGT-----VNCDKANCSTTFCNGGTC-----FY--PGKCI--CPPGLEGE-- 271

Qy      140  TMECVSIHVCVSEWNPSPCTKKGTGKGTETRVRELIQHPSAKGNLCPP----- 193
Db      272  -----QCEISKCP--QPCNNGKCTG---KSKCKSGYGGDLCSKFPVCEPCGGAHG 318

Qy      194  -TNETRKCTVQRKKQKGRKKGRKKK-----KPNKGESKEAIPDSKSLESKEIP 246
Db      319  TCHEPNKC-----QQEGWHGRHCNKRYSALIHALLRPAGALRQHTPSLKAERDRDP 373

Qy      247  E 247
Db      374  E 374
```

Search completed: June 29, 2004, 17:04:01  
Job time : 16.2464 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 17:02:26 ; Search time 36.7755 Seconds

(without alignments)  
2098.641 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISWLFILNMEYI.....QKKRKVDKQKSVSVTVH 273

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1516	100.0	273	10	US-09-894-912A-13
2	1505	99.3	272	10	US-09-894-912A-10
3	1505	99.3	272	10	US-09-894-912A-34
4	1505	99.3	272	14	US-10-125-852-23
5	1472	97.1	265	10	US-09-894-912A-26
6	1472	97.1	292	9	US-09-745-763-166
7	1472	97.1	292	10	US-09-894-912A-48
8	1472	97.1	292	14	US-10-125-852-25
9	1400	92.3	251	10	US-09-894-912A-16
10	1310.5	86.4	279	10	US-09-894-912A-32
11	1275	84.1	225	14	US-10-185-770-4
12	1213	80.0	239	12	US-10-087-192-1782
13	1178	77.7	239	15	US-10-094-886-172
14	1111	73.3	195	15	US-10-094-886-176
15	1107	73.0	195	15	US-10-094-886-174

16	981	64.7	180	12	US-10-087-192-1779	Sequence 1779, Ap
17	903	59.6	160	10	US-09-894-912A-14	Sequence 14, Appl
18	903	59.6	160	12	US-10-276-174-1744	Sequence 1744, Ap
19	656	43.3	263	14	US-10-125-852-18	Sequence 18, Appl
20	648.5	42.8	243	14	US-10-125-852-21	Sequence 21, Appl
21	644	42.5	265	14	US-10-125-852-24	Sequence 24, Appl
22	638	42.1	229	10	US-09-894-912A-25	Sequence 25, Appl
23	626.5	41.3	236	15	US-10-108-260A-4829	Sequence 4829, Ap
24	584.5	38.6	243	14	US-10-185-770-2	Sequence 2, Appl
25	577.5	38.1	243	14	US-10-125-852-13	Sequence 13, Appl
26	577.5	38.1	243	16	US-10-467-042-12	Sequence 12, Appl
27	573	37.8	250	14	US-10-125-852-3	Sequence 3, Appl
28	545	35.9	222	14	US-10-125-852-15	Sequence 15, Appl
29	540.5	35.7	229	14	US-10-125-852-6	Sequence 6, Appl
30	400.5	26.4	190	12	US-10-087-192-873	Sequence 873, App
31	365.5	24.1	161	12	US-10-087-192-876	Sequence 876, App
32	276	18.2	46	10	US-09-894-912A-18	Sequence 18, Appl
33	234.5	15.5	131	14	US-10-125-852-9	Sequence 9, Appl
34	221	14.6	42	10	US-09-894-912A-22	Sequence 22, Appl
35	213	14.1	37	10	US-09-894-912A-20	Sequence 20, Appl
36	202	13.3	110	14	US-10-125-852-11	Sequence 11, Appl
37	186.5	12.3	1548	14	US-10-180-903-2	Sequence 2, Appl
38	177.5	11.7	43	14	US-10-125-852-7	Sequence 7, Appl
39	173.5	11.4	969	10	US-09-961-403-6	Sequence 6, Appl
40	160.5	10.6	337	12	US-10-357-820-20	Sequence 20, Appl
41	160.5	10.6	343	12	US-10-357-820-8	Sequence 8, Appl
42	160.5	10.6	365	12	US-10-357-820-4	Sequence 4, Appl
43	160.5	10.6	365	12	US-10-357-820-6	Sequence 6, Appl
44	160.5	10.6	373	12	US-10-357-820-10	Sequence 10, Appl
45	160.5	10.6	479	9	US-09-764-898-221	Sequence 221, App

ALIGNMENTS

RESULT 1

US-09-894-912A-13  
; Sequence 13, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-894-912A-13

Query Match 100.0%; Score 1516; DB 10; Length 273;  
Best Local Similarity 100.0%; Pred. No. 1.6e-115;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGLRLISWLFILNMEYIGSNASRGRQRMRHPNVSCCGGCATCSDYNGCLSCPK	60
Db	1	MGLRLISWLFILNMEYIGSNASRGRQRMRHPNVSCCGGCATCSDYNGCLSCPK	60
Qy	61	RLPFLALRIGMKIGVCLSSCPGGYGYTRYDINKTKCKADCTCFNKPNCTCKSGFY	120

Db 61 RLFFALERIGHKQIGVCLSSCPGYYGTRYPDINKTKADCDTCFNKNFCTCKSGFY 120  
Qy 121 LHLGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 180  
Db 121 LHLGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 180  
Qy 181 IOHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNKNGESKEAIPDSKSL 240  
Db 181 IOHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNKNGESKEAIPDSKSL 240  
Qy 241 SSKPEORENKQOKKRVQDKOKSVSVTVH 273  
Db 241 SSKPEORENKQOKKRVQDKOKSVSVTVH 273

## RESULT 2

US-09-894-912A-10  
; Sequence 10, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-10

Query Match 99.3%; Score 1505; DB 10; Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.3e-114;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 HRLISWLFILNFMFYGSONASRGRORRHPNVNSQCGGCATCSYNGCLCKPRL 62  
Db 2 HRLISWLFILNFMFYGSONASRGRORRHPNVNSQCGGCATCSYNGCLCKPRL 61  
Qy 63 PFALERIGHKQIGVCLSSCPGYYGTRYPDINKTKADCDTCFNKNFCTCKSGFY 122  
Db 62 PFALERIGHKQIGVCLSSCPGYYGTRYPDINKTKADCDTCFNKNFCTCKSGFY 121  
Qy 123 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 182  
Db 122 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 181  
Qy 183 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNKNGESKEAIPDSKSL 242  
Db 182 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNKNGESKEAIPDSKSL 241  
Qy 243 SSKPEORENKQOKKRVQDKOKSVSVTVH 273  
Db 242 SSKPEORENKQOKKRVQDKOKSVSVTVH 272

## RESULT 3

US-09-894-912A-34  
; Sequence 34, Application US/09894912A

; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 34  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-34

Query Match 99.3%; Score 1505; DB 10; Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.3e-114;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 HRLISWLFILNFMFYGSONASRGRORRHPNVNSQCGGCATCSYNGCLCKPRL 62  
Db 2 HRLISWLFILNFMFYGSONASRGRORRHPNVNSQCGGCATCSYNGCLCKPRL 61  
Qy 63 PFALERIGHKQIGVCLSSCPGYYGTRYPDINKTKADCDTCFNKNFCTCKSGFY 122  
Db 62 PFALERIGHKQIGVCLSSCPGYYGTRYPDINKTKADCDTCFNKNFCTCKSGFY 121  
Qy 123 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 182  
Db 122 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 181  
Qy 183 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNKNGESKEAIPDSKSL 242  
Db 182 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNKNGESKEAIPDSKSL 241  
Qy 243 SSKPEORENKQOKKRVQDKOKSVSVTVH 273  
Db 242 SSKPEORENKQOKKRVQDKOKSVSVTVH 272

## RESULT 4

US-10-125-852-23  
; Sequence 23, Application US/10125852  
; Publication No. US20030032034A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1  
; FILE REFERENCE: HYS-43A  
; CURRENT APPLICATION NUMBER: US/10/125,852  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: US 60/316,368  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 23  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-125-852-23

Query Match 99.3%; Score 1505; DB 14; Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.3e-114;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HURLISWLFILNFMFMEYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 62  
|||  
DB 2 HURLISWLFILNFMFMEYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 61  
|||

QY 63 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYLLH 122  
|||  
DB 62 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYLLH 121  
|||

QY 123 LKCLDNCPEGLNHNTHMECVSIHVECVSEWNPSPCTKKGTCGPKGTETRVREIIQ 182  
|||  
DB 122 LKCLDNCPEGLNHNTHMECVSIHVECVSEWNPSPCTKKGTCGPKGTETRVREIIQ 181  
|||

QY 183 HPSAKGNLCPPTNETRKCTVQRKKCKGGRKKGRKKRKKPKNKGSKKAI PDSKSLSS 242  
|||

DB 182 HPSAKGNLCPPTNETRKCTVQRKKCKGGRKKGRKKRKKPKNKGSKKAI PDSKSLSS 241  
|||

QY 243 KBPEORENKQKKRKVDKQKSVSVTVH 273  
|||  
DB 242 KBPEORENKQKKRKVDKQKSVSVTVH 272  
|||

RESULT 5  
US-09-894-912A-26  
; Sequence 26, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-26

Query Match 97.1%; Score 1472; DB 10; Length 265;  
Best Local Similarity 100.0%; Pred. No. 5.9e-112;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HURLISWLFILNFMFMEYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 62  
|||  
DB 2 HURLISWLFILNFMFMEYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 61  
|||

QY 63 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYLLH 122  
|||  
DB 62 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYLLH 121  
|||

QY 123 LKCLDNCPEGLNHNTHMECVSIHVECVSEWNPSPCTKKGTCGPKGTETRVREIIQ 182  
|||  
DB 122 LKCLDNCPEGLNHNTHMECVSIHVECVSEWNPSPCTKKGTCGPKGTETRVREIIQ 181  
|||

QY 183 HPSAKGNLCPPTNETRKCTVQRKKCKGGRKKGRKKRKKPKNKGSKKAI PDSKSLSS 242  
|||

DB 182 HPSAKGNLCPPTNETRKCTVQRKKCKGGRKKGRKKRKKPKNKGSKKAI PDSKSLSS 241  
|||

QY 243 KBPEORENKQKKRKVDKQKSVSVTVH 273  
|||

DB 242 KBPEORENKQKKRKVDKQKSVSVTVH 272  
|||

RESULT 6  
US-09-745-763-166  
; Sequence 166, Application US/09745763  
; Patent No. US20020065394A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; McCoy, John M.  
; Lavallie, Edward R.  
; Collins-Racie, Lisa A.  
; Evans, Cheryl  
; Merberg, David  
; Treacy, Maurice  
; Spaulding, Vikki  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; ENCODING THEM  
; NUMBER OF SEQUENCES: 219  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/745,763  
; FILING DATE: 18-Jun-2000  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 166:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 292 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 166:  
US-09-745-763-166

Query Match 97.1%; Score 1472; DB 9; Length 292;  
Best Local Similarity 100.0%; Pred. No. 6.6e-112;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HURLISWLFILNFMFMEYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 62  
|||

DB 2 HURLISWLFILNFMFMEYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 61  
|||

QY 63 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYLLH 122  
|||

DB 62 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYLLH 121  
|||

QY 123 LKCLDNCPEGLNHNTHMECVSIHVECVSEWNPSPCTKKGTCGPKGTETRVREIIQ 182  
|||

DB 122 LKCLDNCPEGLNHNTHMECVSIHVECVSEWNPSPCTKKGTCGPKGTETRVREIIQ 181  
|||

QY 183 HPSAKGNLCPPTNETRKCTVQRKKCKGGRKKGRKKRKKPKNKGSKKAI PDSKSLSS 242  
|||

Db 182 HPSAKGNLCPTNTRKCTVQRKKCKQGBRGKKGRKRKRKPKPKNGSKSAIPDSKSLSS 241  
Qy 243 KEIPEORENKQKKKRVQDKQS 266  
Db 242 KEIPEORENKQKKKRVQDKQS 265

RESULT 7  
US-09-894-912A-48  
; Sequence 48, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 48  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-48

Query Match 97.1%; Score 1472; DB 10; Length 292;  
Best Local Similarity 100.0%; Pred. No. 6.6e-112;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 HURLISWLFILNFMEYIGSQNASRRRRRMRHPNVSQCGGCATCSYNGCLCKPRL 62  
Db 2 HURLISWLFILNFMEYIGSQNASRRRRRMRHPNVSQCGGCATCSYNGCLCKPRL 61  
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCDTCFNKNFCTCKKSGFYH 122  
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCDTCFNKNFCTCKKSGFYH 121  
Qy 123 LGKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 182  
Db 122 LGKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 181  
Qy 183 HPSAKGNLCPTNTRKCTVQRKKCKQGBRGKKGRKRKRKPKPKNGSKSAIPDSKSLSS 242  
Db 243 KEIPEORENKQKKKRVQDKQS 266  
Qy 242 KEIPEORENKQKKKRVQDKQS 265

RESULT 8  
US-10-125-852-25  
; Sequence 25, Application US/10125852  
; Publication No. US20030032034A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1  
; FILE REFERENCE: HYS-43A  
; CURRENT APPLICATION NUMBER: US/10/125,852  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-125-852-25

Query Match 97.1%; Score 1472; DB 14; Length 292;  
Best Local Similarity 100.0%; Pred. No. 6.6e-112;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 HURLISWLFILNFMEYIGSQNASRRRRRMRHPNVSQCGGCATCSYNGCLCKPRL 62  
Db 2 HURLISWLFILNFMEYIGSQNASRRRRRMRHPNVSQCGGCATCSYNGCLCKPRL 61  
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCDTCFNKNFCTCKKSGFYH 122  
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCDTCFNKNFCTCKKSGFYH 121  
Qy 123 LGKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 182  
Db 122 LGKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 181  
Qy 183 HPSAKGNLCPTNTRKCTVQRKKCKQGBRGKKGRKRKRKPKPKNGSKSAIPDSKSLSS 242  
Db 182 HPSAKGNLCPTNTRKCTVQRKKCKQGBRGKKGRKRKRKPKPKNGSKSAIPDSKSLSS 241  
Qy 243 KEIPEORENKQKKKRVQDKQS 266  
Db 242 KEIPEORENKQKKKRVQDKQS 265

RESULT 9  
US-09-894-912A-15  
; Sequence 15, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 16  
; LENGTH: 251  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-16

Query Match 92.3%; Score 1400; DB 10; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.1e-106;  
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 23 QNASRRRRRMRHPNVSQCGGCATCSYNGCLCKPRLPFALERIGMKQIGVCLSSCP 82  
Db 1 QNASRRRRRMRHPNVSQCGGCATCSYNGCLCKPRLPFALERIGMKQIGVCLSSCP 60

Qy	83	SGYVGYTRYPDINKCTYKCADCDTCFNKNPCTCKCKSGFYHLGKCLDNCPEGLEANNHTWE	142
Db	61	SGYVGYTRYPDINKCTYKCADCDTCFNKNPCTCKCKSGFYHLGKCLDNCPEGLEANNHTWE	120
Qy	143	CVSIVHCEVSEBWNPWSPCTKKGKTCGFKRGKTETRVREIIQHPSAKGNI <sup>1</sup> CPPTNETRKCTV	202
Db	121	CVSIVHCEVSEBWNPWSPCTKKGKTCGFKRGKTETRVREIIQHPSAKGNI <sup>1</sup> CPPTNETRKCTV	180
Qy	203	QRKCKQGERGCKGRERKRKXPNKGBESKEAIPDSKSLSSSKEI <sup>1</sup> PPQENKQOQKCKKVQD	262
Db	181	QRKCKQGERGCKGRERKRKXPNKGBESKEAIPDSKSLSSSKEI <sup>1</sup> PPQENKQOQKCKKVQD	240
Qy	263	KQKSVSVSTVH	273
Db	241	KQKSVSVSTVH	251

```

RESULT 10
US-09-894-912A-32
; Sequence 32, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 279
; TYPE: prt
; ORGANISM: Mus musculus
US-09-894-912A-32

```

RESULT 11  
US-10-185-770-4

```

; Sequence 4, Application US/10185770
; Publication No. US20030022217A1
; GENERAL INFORMATION:
; APPLICANT: CECCARDI, Toni et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: US$ THERSOP
; FILE REFERENCE: CL0001247
; CURRENT APPLICATION NUMBER: US/10/185,770
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/301,852
; PRIOR FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-185-770-4

Query Match      84.1%; Score 1275; DB 14; Length 225;
Best Local Similarity 100.0%; Pred. No. 5.5e-96;
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3  HLRILSWLFILNPMVEYIGSONASRGRRQRMHPNVISQGGCGGCATCSDYNGCLSCPKRL 62
Db      2  HLRILSWLFILNPMVEYIGSONASRGRRQRMHPNVISQGGCGGCATCSDYNGCLSCPKRL 61

Qy      63  PFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNKNFCTKCKSGGYLH 122
Db      62  PFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNKNFCTKCKSGGYLH 121

Qy      123  LGKCLDNCPLGLEANNHMTVCISIVHCEVSEMPNWPSPCTTKGKTCGFKRGTTETVRRIIQ 182
Db      122  LGKCLDNCPLGLEANNHMTVCISIVHCEVSEMPNWPSPCTTKGKTCGFKRGTTETVRRIIQ 181

Qy      183  HPSAKGNLCPTNTRKCTVORKKCOXGRRGKGRERKQKPNK 226
Db      182  HPSAKGNLCPTNTRKCTVORKKCOXGRRGKGRERKQKPNK 225

```

```

RESULT 12
US-10-087-192-1782
; Sequence 1782, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1782
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-1782

```

Query Match	80.0%	Score 1213	DB 12	Length 239
Best Local Similarity	97.3%	Pred. No. 6.6e-91		
Matches 214	Conservative 1	Mismatches 5	Indels 0	Gaps 0

  

Qy	3	HRLRLISWLFILINPMWEYIGSSQASRGRRRRRHPNVISQCGGCGATCSDYNGCLSKCPRL	62
nb	19	HRLRLISWLFILINPMWEYIGSSQASRGRRRRRHPNVISQCGGCGATCSDYNGCLSKCPRL	78



QY 63 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKADCDTCFNNFCTKCKSGFYLL 122  
DB 79 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKADCDTCFNNFCTKCKSGFYLL 138  
QY 123 LKGLDNCPEGLEANNHTMECVSVHCEVSEWNPSPCTKGGKTCGPKGTETRVREIIQ 182  
DB 139 LKGLDNCPEGLEANNHTMECVSVHCEVSEWNPSPCTKGGKTCGPKGTETRVREIIQ 198  
QY 183 HPSAKGNLCPPNTRKCTVQRKKCKQGGKGRERK 222  
DB 199 HPSAKGNLCPPNTRKCTVQRKKCKQGGKGRGTIIIGEEKK 238

RESULT 13

US-10-094-886-172  
; Sequence 172, Application US/10094886  
; Publication No. US20040002120A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Catherine  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Taupier, Raymond J., Jr.  
; APPLICANT: Miller, Charles  
; APPLICANT: Casman, Stacie  
; APPLICANT: Pena, Carol  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Pochart, Pascal  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Spaderma, Steven  
; APPLICANT: LaRochele, William  
; APPLICANT: Zhong, Mei

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
FILE REFERENCE: 21402-290 B  
CURRENT APPLICATION NUMBER: US/10/094,886  
PRIOR FILING DATE: 2002-03-07  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-08-17  
PRIOR FILING DATE: 2001-05-02  
PRIOR FILING DATE: 2001-05-10  
PRIOR FILING DATE: 2001-09-10  
PRIOR FILING DATE: 2001-08-17  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-08-21  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-03-09  
PRIOR FILING DATE: 2001-06-07  
PRIOR FILING DATE: 2001-06-26  
PRIOR FILING DATE: 2001-08-21  
Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 298  
; SOFTWARE: Patentin 2.1  
; SEQ ID NO 172  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-886-172

Query Match 77.7%; Score 1178; DB 15; Length 239;  
Best Local Similarity 97.7%; Pred. No. 4.6e-88;  
Matches 210; Conservative 0; Mismatches 1; Indels 4; Gaps 2;

QY 3 HURLISWLFIIINFMEYIGSONASRRGRORRMRHPNVSQCGCGCATCSDYNGCLSCKPRLL 62  
DB 2 HURLISWLFIIINFMEYIGSONASRRGRORRMRHPNVSQCGCGCATCSDYNGCLSCKPRLL 61  
QY 63 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCT-KC---KADCDTCFNNFCTKCKSG 118  
DB 62 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTSKPHEKADCDTCFNNFCTKCKSG 121  
QY 119 FYLHLGKCLDNCPEGLEANNHTMECVSVHCEVSEWNPSPCTKGGKTCGPKGTETRV 178  
DB 122 FYLHLGKCLDNCPEGLEANNHTMECVSVHCEVSEWNPSPCTKGGKTCGPKGTETRV 181  
QY 179 EIIQHPSAKGNLCPPNTRKCTVQRKKCKQGGK 213  
DB 182 EIIQHPSAKGNLCPPNTRKCTVQRKKCKQGGK 216

RESULT 14

US-10-094-886-176  
; Sequence 176, Application US/10094886  
; Publication No. US20040002120A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Catherine  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Taupier, Raymond J., Jr.  
; APPLICANT: Miller, Charles  
; APPLICANT: Casman, Stacie  
; APPLICANT: Pena, Carol  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Pochart, Pascal  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Spaderma, Steven  
; APPLICANT: LaRochele, William  
; APPLICANT: Zhong, Mei  
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
FILE REFERENCE: 21402-290 B  
CURRENT APPLICATION NUMBER: US/10/094,886  
CURRENT FILING DATE: 2002-03-07  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-08-17  
PRIOR FILING DATE: 2001-05-02  
PRIOR FILING DATE: 2001-05-10  
PRIOR FILING DATE: 2001-09-10  
PRIOR FILING DATE: 2001-08-17  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-08-21  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-03-08  
PRIOR FILING DATE: 2001-03-09  
PRIOR FILING DATE: 2001-06-07  
PRIOR FILING DATE: 2001-06-26  
PRIOR FILING DATE: 2001-08-21  
Remaining Prior Application data removed - See File Wrapper or PALM.

; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/318,510  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/314,018  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/274,849  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/296,693  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: 60/313,626  
; PRIOR FILING DATE: 2001-08-21  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 298  
; SOFTWARE: PatentIn 2.1  
; SEQ ID NO 176  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-886-176

Query Match 73.3%; Score 1111; DB 15; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.1e-82;  
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 21 GSNASRRRRRQRMHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 80  
DB 1 GSNASRRRRRQRMHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 60  
  
QY 81 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 140  
DB 61 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 120  
  
QY 141 MCVSVIHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 200  
DB 121 MCVSVIHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 180  
  
QY 201 TVQRKKCKQKRG 213  
DB 181 TVQRKKCKQKRG 193

RESULT 15  
US-10-094-886-174  
; Sequence 174, Application US/10094886  
; Publication No. US20040002120A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Catherine  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Taupier, Raymond J., Jr.  
; APPLICANT: Miller, Charles  
; APPLICANT: Casman, Stacie  
; APPLICANT: Pena, Carol  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Gerlach, Valerie

; APPLICANT: Pochart, Pascal  
; APPLICANT: Fernandes, Eima  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Spaderna, Steven  
; APPLICANT: LaRochele, William  
; APPLICANT: Zhong, Wei  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD;  
; FILE REFERENCE: 21402-290 B  
; CURRENT APPLICATION NUMBER: US/10/094,886  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/274,322  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/313,182  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/288,052  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/318,510  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/314,018  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/274,849  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/296,693  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: 60/313,626  
; PRIOR FILING DATE: 2001-08-21  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 298  
; SOFTWARE: PatentIn 2.1  
; SEQ ID NO 174  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-886-174

Query Match 73.0%; Score 1107; DB 15; Length 195;  
Best Local Similarity 99.5%; Pred. No. 2.2e-82;  
Matches 192; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 21 GSNASRRRRRQRMHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 80  
DB 1 GSNASRRRRRQRMHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 60  
  
QY 81 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 140  
DB 61 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 120  
  
QY 141 MCVSVIHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 200  
DB 121 MCVSVIHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 180  
  
QY 201 TVQRKKCKQKRG 213  
DB 181 TVQRKKCKQKRG 193

Search completed: June 29, 2004, 17:15:57  
Job time : 40.7755 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:56:05 ; Search time 13.915 Seconds  
(without alignments)  
1887.186 Million cell updates/sec

Title: US-09-894-912a-13  
Perfect score: 1516  
Sequence: 1 MGLRLISWLFILNFMVY.....QQKRRKVDKQKSVSVTVH 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: Pirl:\*  
2: Pirl:\*  
3: Pirl:\*  
4: Pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	192	12.7	1299	2 T43251	furin (EC 3.4.21.7)
2	186.5	12.3	1548	2 S34583	serine proteinase
3	184	12.1	962	2 JCS571	subtilisin-like pr
4	184	12.1	975	2 JCS570	subtilisin-like pr
5	176.5	11.6	1680	2 A43434	furin (EC 3.4.21.7)
6	173.5	11.4	969	1 A39490	subtilisin-like pr
7	168	11.1	915	2 B48225	probable propotei
8	167.5	11.0	932	2 I52527	PACE4A - mouse (fr
9	167	11.0	915	1 A48225	subtilisin-like pr
10	162	10.7	899	2 G02428	subtilisin-like pr
11	162	10.7	915	2 J6148	subtilisin-like pr
12	161	10.6	440	2 T24232	hypothetical prote
13	158.5	10.5	379	2 I53282	Gene PACE4 protein
14	156.5	10.3	937	2 A47723	P-spondin precursor
15	156	10.3	803	2 A47723	protein bli-4D [im
16	153.5	10.1	942	2 D87803	Notch B protein -
17	150.5	9.9	1203	2 A49175	insulin-like growt
18	149	9.8	1371	2 A33837	Notch protein - Af
19	148.5	9.8	2524	2 A35844	cell-fate determin
20	147.5	9.7	2471	2 A49128	probable kexin (EC
21	145	9.6	570	2 T37314	hypothetical prote
22	145	9.6	1620	2 T27883	P-spondin - rat
23	143.5	9.5	807	2 A38152	Wnt inhibitory fac
24	142	9.4	378	2 B59180	trophozoite cystei
25	142	9.4	677	2 C42125	insulin-like growt
26	139	9.2	1367	1 IGHUR1	notch protein homo
27	139	9.2	2555	2 A40043	protein-tyrosine k
28	138.5	9.1	1369	2 S70713	notch protein homo
29	138	9.1	2531	2 S18188	notch protein homo

30	138	9.1	2531	2 A46019	notch-1 protein -
31	137	9.0	2437	2 S42612	transmembrane prot
32	136	9.0	540	2 B47417	insulin receptor-r
33	135.5	8.9	1382	1 INHUR	insulin receptor p
34	134.5	8.9	1111	2 T26972	hypothetical prote
35	134.5	8.9	1372	2 A34157	insulin receptor p
36	134.5	8.9	1383	2 A36080	insulin receptor p
37	129.5	8.5	2101	2 S57245	insulin receptor (
38	129.5	8.5	2148	1 A56081	insulin receptor -
39	128	8.4	327	2 A46484	apoptosis-mediati
40	128	8.4	1268	2 B36502	insulin receptor-r
41	127	8.4	861	2 A48825	Notch homolog Motc
42	126.5	8.3	3871	2 T22812	hypothetical prote
43	125.5	8.3	667	2 A48579	trophozoite surfac
44	125	8.2	837	2 S43656	furin (EC 3.4.21.7
45	124	8.2	1274	2 T42017	cysteine rich prot

ALIGNMENTS

RESULT 1

T43251  
furin (EC 3.4.21.75) - fall armyworm  
N;Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; serin  
C;Species: Spodoptera frugiperda (fall armyworm)  
C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C;Accession: T43251  
R;Cleplik, M.; Klenk, H.  
submitted to the EMBL Data Library, January 1996  
A;Description: Cloning and functional characterization of FURIN from Spodoptera frugiper  
A;Reference number: Z22368  
A;Accession: T43251  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1299 <CIB>  
A;Cross-references: EMBL:Z68888; NID:g1167859; PID:e219690; PIDN:CAA93116.1  
A;Experimental source: Clone Sfurin 6; ovary  
C;Function:  
A;Description: responsible for the endoproteolytic processing of proproteins with specif  
C;Keywords: hydrolase; serine proteinase

Query Match	12.7%	Score 192;	DB 2;	Length 1299;
Best Local Similarity	27.8%	Pred. No. 2.6e-05;		
Matches	63;	Conservative	26;	Mismatches 78; Indels 60; Gaps 12;
QY	37	NVSQGGCGGCGATCSD--YNGCLSCPKPRIFALERIGMKQIGVCLSSCFSGYGYTRYPDINK 95		
Db	795	SVCRCPCAAHCATCSEADGCTSCHEHL-----VLHDGTCWASCPFSHYET---BDDM 843		
QY	96	CTYCKKADCTCF--NKNFTCKSGGFYLHGLKLDNCPGLEHANNHTMTCVSIYHCEVSE 153		
Db	844	CARHESCDTCQGGTGTCTCHPSTPSTYALDGRCVTSCPPAYADKKRKC---MRCPVG- 899		
QY	154	WNWSPCTKGGKTCGFRGTETVRETIQHPSAKGNLCPP---TNETKRC-TVQKKCKOK 209		
Db	900	---CSTCT-----SAPCLSCBPFWELNKKKGKCMVSGDKCSA 933		
QY	210	GERGKGRERKRKKPNK-----GESKE---AIPDSKSLSSKEIPE 247		
Db	934	GBFAV---DQKCRKCPACDSCYGENEGHCLTCFNPNNLLQDYKCVPE 977		

RESULT 2

S34583  
serine proteinase (EC 3.4.21.-) PC6B - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999  
C;Accession: S34583  
R;Nakagawa, T.; Murakami, K.; Nakayama, K.  
FEBS Lett. 327, 165-171, 1993  
A;Title: Identification of an isoform with an extremely large Cys-rich region of PC6, a  
A;Reference number: S34583; MUID:93327934; PMID:8335106

A:Accession: S34583  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1548 <NAK>  
A:Cross-references: GB:D17583; NID:G407344; PIDN:BAA04507.1; PID:dl005033; PID:G440374  
C:Keywords: hydrolase; serine proteinase

Query Match 12.1%; Score 186.5; DB 2; Length 1548;  
Best Local Similarity 25.6%; Pred. No. 6.8e-05;  
Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;

QY 39 SQCGGCGATCSDYNG--CLSKPRFLPFFALERIGMKQIGVCLSSCPGYYGTRYPI--N 94  
DB 880 NQCHSSKTC---NSGLASCPTGYLWLO-----ACVSCQCG-----TWPSVTSG 924  
QY 95 KCTKCKACDCTCNKPNCTKCKS----GFYLHLGKCLDNCPEGLAEANNHTME--CVSIVHC 149  
DB 925 SCEKSEDCVSCGADLCQQLSQPDNTLLHGRGCVHSCPGFYAKDGVCEHCSS--PC 982  
QY 150 EVSEWNPSPCTKKG-----KTC-----GPKRGTRVREIILQHPSAKGNL- 190  
DB 983 KTCGNATSCNSCEGDFVLHGVCMKTCPEKHVAVGVCKPCPCQDCIHEKTKCKCMP 1042  
QY 191 -----CPPT--NETRKTCTVQRKKCKQKGERGKGRERKRPKPKNKGSKKAIPDS 236  
DB 1043 DFLYNDMCHRSKPCXFPYDMRQCVPCHKNC-----LBCNGPKEDDCKVCADTS 1091  
QY 237 KSLSS---KEIPE-----QRENKQ 253  
DB 1092 KALLHNGCLDCECPGTYKBEENDS 1115

## RESULT 3

JCS571  
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-II -  
C:Species: Homo sapiens (man)  
C:Date: 23-Sep-1997 #sequence\_revision 23-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: JCS571  
R:Morii, K.; Kii, S.; Teuji, A.; Nagahama, M.; Imanaki, A.; Hayashi, K.; Akamatsu, T.; Na  
J. Biochem. 121, 941-948, 1997  
A:Title: A novel human PACE4 isoform, PACE4B is an active processing protease containing  
A:Reference number: JCS570; MUID:97335942; PMID:9192737  
A:Accession: JCS571  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-962 <MOR>  
A:Cross-references: DDBJ:D87994; NID:G2330550; PIDN:BAA21792.1; PID:G2330551  
A:Experimental source: brain cerebellum  
C:Comment: This enzyme is a processing protease and responsible for processing of various  
ch it is retained intracellularly.

C:Genetics:  
A:Gene: GDB:PACE4  
A:Cross-references: GDB:131390; OMIM:167405  
A:Map position: 15q26-15q26  
A:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase  
F:1-62/Domain: signal sequence #status predicted <SIG>  
F:63-149/Domain: propeptide #status predicted <PRO>  
F:196-434/Domain: subtilisin homology <SBT>  
F:938-954/Domain: hydrophobic cluster #status predicted <HCL>  
F:205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted  
F:259/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.1%; Score 184; DB 2; Length 962;  
Best Local Similarity 26.6%; Pred. No. 7e-05;  
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;  
QY 19 YIGSQNASRGRQRMRHNPVSGCGGCATCSD--YNGCLSKPRFLPFFALERIGMKQIGV 76  
DB 724 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGFI-----HHQEMNT 766

QY 77 CLSSCPGYYGTRYPDINKTKKADCTCFNK-NFTCKSGFYHLGKCLDNCPEGLE 135  
Query Match 12.1%; Score 184; DB 2; Length 962;  
Best Local Similarity 26.6%; Pred. No. 7e-05;  
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;  
QY 19 YIGSQNASRGRQRMRHNPVSGCGGCATCSD--YNGCLSKPRFLPFFALERIGMKQIGV 76  
DB 724 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGFI-----HHQEMNT 766

DB 767 CVTLCPAGFYADE--SQKNCLKCHPCKKCVDPBEKCTVCKBGFSLARGSCIPDCBGTY 824  
QY 136 ANNHTMEC-----VSIVHCEVS--EMNPSPCTKKGTCGPKRGTRVRE 179  
DB 825 FDSLIROGECCHTCTGTCVGPGRBECIHCAKNPHFHDW-----KCVPAKGFGFYPE----- 875  
QY 180 IIQHPSAKGNLC-----PPTNETRKTCTVQRKKCKQKGERGKGRERKRPKPKNKG 228  
DB 876 --EMPGLPHKVCRRYCPGGB-RQATVS-----SKGVPG--GQSLSSASSPAGE 919

## RESULT 4

JCS570  
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-I -  
C:Species: Homo sapiens (man)  
C:Date: 23-Sep-1997 #sequence\_revision 23-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: JCS570  
R:Morii, K.; Kii, S.; Teuji, A.; Nagahama, M.; Imanaki, A.; Hayashi, K.; Akamatsu, T.; Na  
J. Biochem. 121, 941-948, 1997  
A:Title: A novel human PACE4 isoform, PACE4B is an active processing protease containing  
A:Reference number: JCS570; MUID:97335942; PMID:9192737  
A:Accession: JCS570  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-975 <MOR>  
A:Cross-references: DDBJ:D87993; NID:G2330548; PIDN:BAA21791.1; PID:G2330549  
A:Experimental source: brain cerebellum  
C:Comment: This enzyme is a processing protease and responsible for processing of various  
ch it is retained intracellularly.

C:Genetics:  
A:Gene: GDB:PACE4  
A:Cross-references: GDB:131390; OMIM:167405  
A:Map position: 15q26-15q26  
A:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: alternative splicing; glycoprotein; hydrolase; serine proteinase  
F:1-62/Domain: signal sequence #status predicted <SIG>  
F:63-149/Domain: propeptide #status predicted <PRO>  
F:196-434/Domain: subtilisin homology <SBT>  
F:952-968/Domain: hydrophobic cluster #status predicted <HCL>  
F:205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted  
F:259/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.1%; Score 184; DB 2; Length 975;  
Best Local Similarity 26.6%; Pred. No. 7.1e-05;  
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;  
QY 19 YIGSQNASRGRQRMRHNPVSGCGGCATCSD--YNGCLSKPRFLPFFALERIGMKQIGV 76  
DB 737 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGFI-----HHQEMNT 779

QY 77 CLSSCPGYYGTRYPDINKTKKADCTCFNK-NFTCKSGFYHLGKCLDNCPEGLE 135  
DB 780 CVTLCPAGFYADE--SQKNCLKCHPCKKCVDPBEKCTVCKBGFSLARGSCIPDCBGTY 837  
QY 136 ANNHTMEC-----VSIVHCEVS--EMNPSPCTKKGTCGPKRGTRVRE 179  
DB 838 FDSLIROGECCHTCTGTCVGPGRBECIHCAKNPHFHDW-----KCVPAKGFGFYPE----- 888  
QY 180 IIQHPSAKGNLC-----PPTNETRKTCTVQRKKCKQKGERGKGRERKRPKPKNKG 228  
DB 889 --EMPGLPHKVCRRYCPGGB-RQATVS-----SKGVPG--GQSLSSASSPAGE 932

## RESULT 5

A43434  
furin (EC 3.4.21.75) 2 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Mar-2000  
C:Accession: A43434  
R:Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dumke, U.; Rentrop, M.; Gateff, B.  
J. Biol. Chem. 267, 17208-17215, 1992  
A:Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc  
A:Reference number: A43434; MUID:92381036; PMID:1512259

A:Accession: A43434  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1680 <ROE>  
A:Cross-references: GB:M94375; NID:g157461; PID:g157462  
A:Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBIP:111934)  
C:Genetics:  
A:Gene: FlyBase:Pur2  
A:Cross-references: FlyBase:FBgn0004598  
C:Superfamily: subtilisin homology  
C:Keywords: hydrolase; serine proteinase; transmembrane protein  
F:409-652/Domain: subtilisin homology <SBT>  
F:418,457,638/Active site: Asp, His, Ser #status predicted

Query Match 11.6%; Score 176.5; DB 2; Length 1680;  
Best Local Similarity 28.0%; Pred. No. 0.00033;  
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

QY 3 HLRILSWLFIILNPMVEYIGSQASRGRRRMRHNPVNSQGGGATCSDY-NGCLSCKPR 61  
DB 1038 HLHVID-LAVCLQPCPGYFENS-----RNTCTVP-----CEPNCASQDHPYCTSCDHH 1087  
QY 62 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKKADCDTCF--NKNFCTKCKSGF 119  
DB 1088 LVMHEHK-----CYSACPLDYET--EDNKCACFCHSTCATCNGPTDQDCITCSSR 1136  
QY 120 YLHLGKCLDNCPEGLEANNHMECVSIHVCEVSEWNPSPCTKGGKTCGFKRGTRTV-R 178  
DB 1137 YAWQNKCLISCPDGFYADKRLCEM-----PCQEGKTC-----TSNGVCS 1177  
QY 179 EIIQHPSAKGNLCPTTNETRKTQVKK-CQKGE 211  
DB 1178 ECLQNT-----LNRDKCIVSGEGGSESE 1203

RESULT 6  
A39490  
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form A - human  
N:Alternate names: kexin homology  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Mar-2000  
C:Accession: A39490  
R:Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landeberg, K.E.; Saltman, D.; Barr, P.J.  
DNA Cell Biol. 10, 757-769, 1991  
A:Title: Identification of a second human subtilisin-like protease gene in the fes/fps region  
A:Reference number: A39490; MUID:92075167; PMID:1741956  
A:Molecule type: mRNA  
A:Residues: 1-969 <KIE>  
A:Cross-references: GB:M80482; NID:g189531; PIDN:AAAS9998.1; PID:g189532  
C:Genetics:  
A:Gene: GDB:PACE4  
A:Cross-references: GDB:131390; OMIM:167405  
A:Map position: 15q26-15q26  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: alternative splicing; hydrolase; serine proteinase  
F:150-969/Product: serine proteinase PACE4 #status predicted <SIG>  
F:196-434/Domain: subtilisin homology <SBT>  
F:205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 11.4%; Score 173.5; DB 1; Length 969;  
Best Local Similarity 28.8%; Pred. No. 0.00034;  
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 19 YIGSQASRGRRRMRHNPVNSQGGGATCSD--YNGCLSCKPRLPFALERIGMKQIGV 76  
DB 737 YFGPTAARRCR-----CHKGETSSRAATQCLSCR-RGFY-----HHQENMT 779  
QY 77 CLSCSPSGYGYTRYPDINKTKKADCDTCFNK-NFTCKSGSYLHLGKCLDNCPEGLE 135  
DB 780 CVTLCPAGFYADB--SQRNCLKHPCKCKVDEPEKTCVCKEGFSLARGSCIPDCPGTY 837  
QY 136 ANNHITMECVSIHVCEVSEWNPSPCTKKGKTCG 168

DB 838 PDSELIRGCGCHH-----TCG 853  
RESULT 7  
B48225  
probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat  
N:Alternate names: PCS precursor  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 23-Feb-1997  
C:Accession: B48225  
R:Lueson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993  
A:Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate proprotein  
A:Reference number: A48225; MUID:93342056; PMID:8341687  
A:Accession: B48225  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-915 <LUS>  
A:Cross-references: GB:L14933  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>  
F:117-915/Product: probable proprotein convertase 5 #status experimental <MAT>  
F:164-402/Domain: subtilisin homology <SBT>  
F:173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.1%; Score 168; DB 2; Length 915;  
Best Local Similarity 26.3%; Pred. No. 0.00075;  
Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;

QY 36 PNVSQ-GCOG-----GCATCDYNGCLSCKPRLPFALERIGMKQIGVCLSSCPGSGYGYTRY 90  
DB 640 PRCSEVCGDGPDPDHTCDLHYHLYKLNNT------ICVSSCPGPHF---H 682  
QY 91 PDINKTKKADCDTCFNKFN--CTKCKSGFYLH--LGKCLDNCPEGLEANNHMECVSI 146  
DB 683 ADKRCCKCAPNCSCFSGSHADQCLSKYGYFLNEETSSCVAQCEGSGYQDKKNIC--- 739  
QY 147 VHCVEWNPSPCTKKGKTC-GFKRGTRTVREIIQHPASAGNLCPTTNETRKTQVOK 205  
DB 740 -----GKSCENKCTCTGFHNCTE-----CKGGL---SLQGSRCVS--- 771  
QY 206 KQKRG 211  
DB 772 TCSDGQ 777

RESULT 8  
I52527  
PACE4A - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999  
C:Accession: I52527  
R:Hosaka, M.; Murakami, K.; Nakayama, K.  
Biomed. Res. 15, 383-390, 1994  
A:Title: PACE4A is a ubiquitous endoprotease that has similar but not identical substrates  
A:Reference number: I52527  
A:Accession: I52527  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-932 <RES>  
A:Cross-references: GB:D50060; NID:g769700; PIDN:BA08777.1; PID:g769701  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
F:172-410/Domain: subtilisin homology <SBT>

Query Match 11.0%; Score 167.5; DB 2; Length 932;  
Best Local Similarity 24.1%; Pred. No. 0.00082;  
Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

QY 19 YIGSQASRGRRRMRHNPVNSQGGGATCSDYN--GCLSKPRLPFALERIGMKQIGV 76



```

A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-379 <BS1>
A,Cross-references: GB:AF122922; NID:G4585369; PIDN:AAD25402.1; PID:G4585370

Query Match      10.5%; Score 158.5; DB 2; Length 379;
Best Local Similarity 24.9%; Pred. NO. 0.0016;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16

QY      42  CQGGCAT---CSDYNGC-----LSCKPRLFFALLERIGMQIGVCLSS-----C 81
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      182  CPGCGKNGSFCNERRICEPDGPHGHCERKALCTPRONG-----GLCVTPGFCIC 232

QY      82  PSGYITRYPDINKTKYC-KADCO-TCFNKNFTCKSGGYHLHGKCLDNCNCPGLEANNH 139
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      233  PGFYG-----VNCCKANKSTTCFNGGTC-----FY--PGKCI--CPPGLEGB-- 271

QY      140  TMBCVSIVHCSEVSEMPNPSPCTKKKGTGCFKRGTRTETVRBIIQHPSSAKGNLCPP----- 193

```

## 2

```

      QY   194 -TNETRKCTVQRKKCKGKGRERKRK-----KPNKGESKEAIPDSKSLSSSKEIP 244
           :|::||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
      Db    319 TCHPEPKC-----QCQEGWHGRHCNRRYEASLTHAIRPAGQAQLRQHTPSLKKAERRDPP 373

      QY       247 E 247
           |
      Db       374 E 374

RESULT 14
IS3282
gene PACB4 protein - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 02-Aug-1996 #sequence_revision 02-Aug-1996
C.Accession: IS3282
R.Johnson, R.C.; Darlington, D.N.; Hand, T.A.; Bloomquist, B.T.; Mains, R.E.
Endocrinology 135, 1178-1185, 1994
A.Title: PACB4: a subtilisin-like endoprotease prevalent in the anterior pituitary gland of the rat.
A.Reference number: IS3282; PMID:8070361
```

A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-937 <RES>  
A:Cross-references: GB:L18994; NID:g946221; PIDN:AAA61987.1; PID:g946222  
C:Genetics:  
A:Gene: PACE4  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
F:177-415/Domains: subtilisin homology <SBT>

Query Match 10.3%; Score 156.5; DB 2; Length 937;  
Best Local Similarity 23.6%; Pred. No. 0.0043;  
Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11

QY 19 YIGSQNASRRGRQRMRHNPVVSQCQGCGATCSYN--GCLSCKRLPFALEIRGMKGIVG 76  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

QY

```

Db      806  FDSELI+RGCECHT+CR+TCVPSREECIHCAKSFHPQDMKVC+PACGB+-----GP----- 853
Qy      177  VREIIQHPSAKNLCPTTETTRKCTVQRKKCQKGR 212
Db      854  -----YPBEMPGL--PHKVCRCDCNCLSCGSSR 881

RESULT 15
A47723

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:24 ; Search time 8.28277 Seconds  
(without alignments)  
1716.232 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISWLFILMFMEYI.....QQKKRKVQDKQSVSVTH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186.5	12.3	1877	1	PCK5 MOUSE
2	184.5	12.2	1696	1	PCK5 BRACL
3	176.5	11.6	1679	1	FUR2 DROME
4	173.5	11.4	969	1	PAC4 HUMAN
5	168	11.1	1877	1	PCK5 RAT
6	162	10.7	913	1	PCK5 HUMAN
7	158.5	10.5	379	1	WIFI HUMAN
8	157.5	10.4	374	1	WIFI XENLA
9	156.5	10.3	937	1	PAC4 RAT
10	156	10.3	803	1	FSPQ XENLA
11	150.5	9.9	2470	1	NTC2 MOUSE
12	149	9.8	1370	1	IGIR RAT
13	148.5	9.8	2524	1	NOTC XENLA
14	147.5	9.7	2471	1	NTC2 RAT
15	145	9.6	943	1	BL14 CAEL
16	143.5	9.5	379	1	WIFI MOUSE
17	143.5	9.5	807	1	FSPQ RAT
18	142	9.4	378	1	WIFI BRARE
19	142	9.4	1373	1	IGIR MOUSE
20	139	9.2	1367	1	IGIR HUMAN
21	139	9.2	2556	1	NTC1 HUMAN
22	138	9.1	2531	1	NTC1 MOUSE
23	138	9.1	2531	1	NTC1 RAT
24	137	9.0	2437	1	NTC1 BRARE
25	135.5	8.9	1382	1	INSR HUMAN
26	134.5	8.9	1372	1	INSR MOUSE
27	134.5	8.9	1383	1	INSR RAT
28	133.5	8.8	2482	1	VWF FIG
29	130.5	8.6	1300	1	IGIR MOUSE
30	130	8.6	2471	1	NTC2 HUMAN
31	129.5	8.5	2146	1	INSR DROME
32	128	8.4	327	1	TNR6 MOUSE
33	128	8.4	1297	1	IGIR HUMAN

34	126	8.3	687	1	VS41 GIALA
35	125.5	8.3	667	1	TS11_GIALA
36	123	8.1	461	1	TRIA_PIG
37	123	8.1	559	1	TRAP PLAPA
38	123	8.1	581	1	IRR RAT
39	119.5	7.9	1429	1	LI12 CAEL
40	119	7.8	1300	1	IRR CAPO
41	119	7.8	2703	1	NOTC DROME
42	118.5	7.8	1477	1	HTK7_HYDAT
43	118	7.8	867	1	SSPO BOVIN
44	117.5	7.8	833	1	DL DROME
45	117.5	7.8	1367	1	LT23 CAEL

RESULT 1  
PCK5\_MOUSE STANDARD; PRT; 1877 AA.  
AC Q04552; Q62040;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PC5) (PC6)  
DE (Subtilisin-like proprotein convertase 6) (SPC6).  
GN PCSK5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Rutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).  
RC STRAIN=ICR; TISSUE=Intestine;  
RX MEDLINE=93327934; PubMed=8335106;  
RA Nakagawa T., Murakami K., Nakayama K.;  
RT "Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";  
RL FEBS Lett. 327:165-171(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
RC TISSUE=Brain, and Intestine;  
RX MEDLINE=93224489; PubMed=8468318;  
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;  
RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACB4.";  
RL J. Biochem. 113:132-135(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
RC TISSUE=Adrenal cortex;  
RX MEDLINE=93342056; PubMed=8341687;  
RA Luessen J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
RN [4]  
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
RX MEDLINE=97103178; PubMed=8947550;  
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;  
RT "The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";  
RL J. Cell Biol. 135:1261-1275(1996).  
RN [5]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=96293359; PubMed=8698813;  
RA Constam D.B., Calton M., Robertson E.J.;  
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

ALIGNMENTS

RL J. Cell Biol. 134:181-191(1996).  
 RP [6]  
 RP DEVELOPMENTAL EXPRESSION.  
 RA MEDLINE=97436919; PubMed=9291583;  
 RX Rancourt S.L., Rancourt D.B.;  
 RT "Murine subtilisin-like proteinase SP6 is expressed during embryonic  
 RT implantation, somitogenesis, and skeletal formation.";  
 RL Dev. Genet. 21:75-81(1997).  
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity  
 CC within the constitutive and regulated secretory pathway. Capable  
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible  
 CC for the maturation of gastrointestinal peptides. May be involved  
 CC in the cellular proliferation of adrenal cortex via the activation  
 CC of growth factors.  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -!- SUBCELLULAR LOCATION: PCSA is SECRETED THROUGH THE REGULATED  
 CC SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO  
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH  
 CC EARLY ENDOSOMES.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=PCSB; Synonyms=Long;  
 CC IsoId=Q04592-1; Sequences=Displayed;  
 CC Name=PCSA; Synonyms=Short;  
 CC IsoId=Q04592-2; Sequences=VSP\_005439;  
 CC -!- TISSUE SPECIFICITY: PCSA is expressed in most tissues but is most  
 CC abundant in the intestine and adrenals. PCSB is expressed in the  
 CC intestine, adrenals and lung but not in the brain.  
 CC -!- DEVELOPMENTAL STAGE: Weakly expressed throughout the embryo,  
 CC except in the developing nervous system, the ribs and the liver,  
 CC but markedly upregulated at discrete sites during development. At  
 CC E6.5, prominent expression observed in differentiated decidua. At  
 CC E7.5, intense expression in extraembryonic endoderm, amnion and  
 CC nascent mesoderm. At E8.5, abundant expression in somites and yolk  
 CC sac followed by a confinement to dermamyotome compartment. Between  
 CC E9.5 and E11.5, abundant expression in AER (thickened ectodermal  
 CC cells of limb buds). At E12.5, expression in the limbs is confined  
 CC to the condensing mesenchym surrounding the cartilage. At this  
 CC stage, strong expression also detected in vertebral and facial  
 CC cartilage primordia and in the muscle of the tongue. At E16.5,  
 CC abundant expression in epithelial cells of the intestinal villi.  
 CC Isoform A is most abundant at all stages but significant levels of  
 CC isoform B occur at E12.5.  
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
 CC assisting the folding of the zymogen within the endoplasmic  
 CC reticulum.  
 CC -!- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain  
 CC sorting information. AC 1 directs TGN localization and interacts  
 CC with the TGN sorting protein PACS-1.  
 CC -!- SIMILARITY: Belongs to peptidase family S8.  
 CC -!- SIMILARITY: Contains 1 homo B/P domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D17583; BAA04507.1; -;  
 DR EMBL; D12619; BAA02143.1; -;  
 DR EMBL; L14932; AAA74636.1; -;  
 DR PIR; A48225; A48225.  
 DR PIR; S34583; S34583.  
 DR HSSP; Q99405; 1MPT.  
 DR MEROPS; S08.076; -;  
 DR MGD; MGI:97515; Pcsks.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.

DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR002884; Peptidase\_S8B.  
 DR InterPro; IPR009020; Protease\_inhib.  
 DR Pfam; PF01483; P\_protease\_1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR ProDom; PD000717; p\_domain; 1.  
 DR SMART; SM00261; FU\_22.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing;  
 KW Transmembrane.  
 FT SIGNAL 1 34  
 FT PROPEP 35 116  
 FT CHAIN 117 1877  
 FT  
 FT DOMAIN 117 1768 PROTEIN CONVERTASE SUBTILISIN/KEXIN  
 FT TRANSMEM 1769 1769 TYPE 5.  
 FT DOMAIN 1790 1877 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 1790 1877 POTENTIAL.  
 FT DOMAIN 117 452 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 464 602 CATALYTIC.  
 FT DOMAIN 638 1753 HOMO B.  
 FT DOMAIN 1825 1844 CYS-RICH MOTIF (CRM) REGION.  
 FT DOMAIN 1856 1877 AC 1.  
 FT SITE 116 117 AC 2.  
 FT SITE 521 523 CLEAVAGE (AUTO-) (BY SIMILARITY).  
 FT ACT SITE 173 173 CELL ATTACHMENT SITE (POTENTIAL).  
 FT ACT SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 383 383 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 667 667 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 754 754 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 804 804 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 854 854 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 951 951 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 1220 1220 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 1317 1317 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 1711 1711 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT CARBOHYD 1733 1733 N-LINKED (GLCNAC..) (POTENTIAL).  
 FT VARSPPLIC 878 915 GEYIDQGHCTCEASCKWGPQEDCISCPVTRVLD ->  
 FT ATESWAEGGFCMLVKNNLCQRYLQQLCKCTPQG  
 FT (in isoform PCSA).  
 FT /FTId=VSP\_005438.  
 FT Missing (in isoform PCSA).  
 FT /FTId=VSP\_005439.  
 FT  
 FT VARSPPLIC 916 1877  
 FT  
 FT SEQUENCE 1877 AA; 209287 MW; EC850E2D20E8A1C3 CRC64;  
 Query Match 12.3%; Score 186.5; DB 1; Length 1877;  
 Best Local Similarity 25.0%; Pred. No. 6.5e-06;  
 Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;  
 QY 39 SQCCQCGCATCSYNG--CLSCXPRLPFALLERIGMKRQIGVCLSCPSGVYTRYPD1--N 94  
 Db 1209 NQPCSSCKTC---NGSLCASCPTGMYLWLO-----ACVSPCPOG----TWFSVTSG 1253  
 QY 95 KCTKCKADCDTCFNKNPCTKCKS---GFYLHLGKCLDNCPEGLANNHTWE-CVSIVHC 149  
 Db 1254 SCCKCEDVCSCGADLCQQLSQPDNTLLHGRGVHSCBPGPYAKGVGVCHECSS--PC 1311  
 QY 150 EVSENPMPSCTKKG-----KTC-----GFKRGTRTVRHIIQHPAKGNL- 190  
 Db 1312 KTCGNATSCNSCBGDFVLHDGVCMWKTPCKHVAVEGVCKHCPERCQDCIHEKCKECP 1371  
 QY 191 -----CPPT--NETRKTQVQRKKQCKGRCKGRKRKKPKNKGESKEAIPDS 236  
 Db 1372 DFFLYNDMCHRSCKPKSFYPMRQCPCPKNC-----LECNKPKEDDCKVCA DTS 1420  
 QY 237 KSLESS---KEIPE-----QRENKQ 253

Db 1421 KALHNGLCLEDCPEGTGYKXENDE 1444

RESULT 2

PK5\_BRACL STANDARD; PRT; 1696 AA.

ID PK5\_BRACL STANDARD; PRT; 1696 AA.

AC Q9NJ15; Q9NJ14; Q9NJ16; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)

DE (proprotein convertase PC6-like) (aPC6).

GN PC6.

OS Branchiostoma californiensis (California lancelet) (Amphioxus).

OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.

OX NCBI\_TaxID=7738;

[1]

SEQUENCE FROM N.A. (ISOFORMS A; B AND C).

RA MEDLINE=20175281; PubMed=10708868;

RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;

RT "Evolution of the prohormone convertases: identification of a homologue of PC6 in the protochordate amphioxus.";

RL Mol. Biol. Evol. 14:338-348(1997).

CC -1- FUNCTION: Likely to represent a widespread endoprotease activity within the constitutive and regulated secretory pathway. Capable of cleavage at the RX(K/R)R consensus motif (By similarity).

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys.

CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.

CC -1- ISOFORM B IS A TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Name=B;

CC IsoId=Q9NJ15-1; Sequence=Displayed;

CC Name=A;

CC IsoId=Q9NJ15-2; Sequence=VSP\_005444, VSP\_005445;

CC Name=C;

CC IsoId=Q9NJ15-3; Sequence=VSP\_005442, VSP\_005443;

CC -1- DOMAIN: The propeptide domain acts as an intramolecular chaperone assisting the folding of the zymogen within the endoplasmic reticulum.

CC -1- SIMILARITY: Belongs to peptidase family S8.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

-----

EMBL; AF184615; AAP26300.1; .

DR EMBL; AF184616; AAP26301.1; .

DR EMBL; AF184817; AAP26302.1; .

DR HSSP; Q99405; IMPT.

DR MEROPS; S08.U08; .

DR InterPro; IPR006212; Furin repeat.

DR InterPro; IPR009030; Grow fac recep.

DR InterPro; IPR000209; Peptidase\_S8.

DR InterPro; IPR002884; Peptidase\_S8B.

DR InterPro; IPR009020; Protease\_inhib.

DR Pfam; PF01483; P\_protease1n; 1.

DR Pfam; PF00082; Peptidase\_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P\_domain; 1.

DR SMART; SM00261; FU; 17.

DR PROSITE; PS00136; SUBTILASE ASP; FALSE\_NEG.

DR PROSITE; PS00137; SUBTILASE HIS; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

Db 1421 KALHNGLCLEDCPEGTGYKXENDE 1444

RESULT 2

PK5\_BRACL STANDARD; PRT; 1696 AA.

ID PK5\_BRACL STANDARD; PRT; 1696 AA.

AC Q9NJ15; Q9NJ14; Q9NJ16; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Proprietary protease subtilisin/kexin type 5 precursor (EC 3.4.21.-) (proprotein convertase PC6-like) (apC6).

GN PC6.

OS Brachiostoma californiensis (California lancelet) (Amphioxus).

OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Brachiostoma.

OX NCBI\_TaxID=7738;

[1]

SEQUENCE FROM N.A. (ISOFORMS A; B AND C).

RA MEDLINE=20175281; PubMed=10708868;

RR Oliva A.A. Jr., Chan S.J., Steiner D.F.;

RT "Evolution of the prothormone convertases: identification of a homologue of PC6 in the protochordate amphioxus.";

RL Mol. Cell. Biochem. 147:338-348(2000).

CC -1- FUNCTION: Likely to represent a widespread endoprotease activity within the constitutive and regulated secretory pathway. Capable of cleavage at the RX(K/R)R consensus motif (By similarity).

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys.

CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.

CC -1- ISOFORM B IS A TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Name=B;

CC IsoId=Q9NJ15-1; Sequence=Displayed;

CC Name=A;

CC IsoId=Q9NJ15-2; Sequence=VSP\_005444, VSP\_005445;

CC Name=C;

CC IsoId=Q9NJ15-3; Sequence=VSP\_005442, VSP\_005443;

CC -1- DOMAIN: The propeptide domain acts as an intramolecular chaperone assisting the folding of the zymogen within the endoplasmic reticulum.

CC -1- SIMILARITY: Belongs to peptidase family S8.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; AF184615; AAP26300.1; -

DR EMBL; AF184616; AAP26301.1; -

DR EMBL; AF184817; AAP26302.1; -

DR HSSP; Q99405; IMPT.

DR MEROPS; S08.U08; -

DR InterPro; IPR006212; Furin repeat.

DR InterPro; IPR009030; Grow fac recep.

DR InterPro; IPR000209; Peptidase\_S8.

DR InterPro; IPR002884; Peptidase\_S8B.

DR InterPro; IPR009020; Protease\_inhib.

DR Pfam; PF01483; P\_protease1n; 1.

DR Pfam; PF00082; Peptidase\_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P\_domain; 1.

DR SMART; SM00261; FU; 17.

DR PROSITE; PS00136; SUBTILASE ASP; FALSE\_NEG.

DR PROSITE; PS00137; SUBTILASE HIS; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

DR

OC Ephydroidea; Drosophilidae; Drosophila.  
 RN NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R, Tuebingen, and Iso-1;  
 RX MEDLINE=92381036; PubMed=1512259;  
 RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,  
 RA Rentrop M., Gateff E.A.P., Leunissen J.A.M., van de Ven W.J.M.;  
 RT "Cloning and functional expression of Dfurin2, a subtilisin-like  
 RT protease processing enzyme of Drosophila melanogaster with multiple  
 RT repeats of a cysteine motif.";  
 RL J. Biol. Chem. 267:17208-17215(1992).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Iso-1;  
 RX MEDLINE=95186060; PubMed=7880443;  
 RA Roebroek A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,  
 RA van de Ven W.J.M.;  
 RT "The Dfur2 gene of Drosophila melanogaster: genetic organization,  
 RT expression during embryogenesis, and pro-protein processing activity  
 RT of its translational product Dfurin2.";  
 RL DNA Cell Biol. 14:223-234(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Finkler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -!- FUNCTION: Furin is likely to represent the ubiquitous endoprotease  
 CC activity within constitutive secretory pathways and capable of  
 CC cleavage at the RX(K/R)R consensus motif (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,  
 CC complement component C3 and von Willebrand factor from their  
 CC respective precursors.  
 CC -!- TISSUE SPECIFICITY: Transient expression in a subset of central

CC nervous system neurons during embryonic stages 12-13. Expression  
 CC in developing tracheal tree from stage 13 to end of embryonic  
 CC development.  
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.  
 CC -!- SIMILARITY: Belongs to peptidase family S9. Furin subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M94375; AAA28551.1; -;  
 DR EMBL; L33831; AAA69860.1; -;  
 DR EMBL; AE003502; AAF48598.1; -;  
 DR PIR; A43434; A43434.  
 DR HSSP; Q99405; LMPT.  
 DR MEROPS; S08\_049; -;  
 DR FlyBase; FBgn0004598; Fur2.  
 DR GO; GO:0004276; F: furin activity; IDA.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR00209; Peptidase\_S8.  
 DR InterPro; IPR002884; Peptidase\_S8B.  
 DR InterPro; IPR009020; Protease\_inhib.  
 DR Pfam; PF01483; P\_protein; 2.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR ProDom; PD000717; P\_domain; 1.  
 DR SMART; SM00261; FU\_10.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;  
 KW Multigene family; Zymogen; Repeat.  
 FT SIGNAL; 1; ? POTENTIAL.  
 FT PROPEP; 318 POTENTIAL.  
 FT CHAIN; 319 1679 FURIN-LIKE PROTEASE 2.  
 FT ACT\_SITE; 417 417 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE; 456 456 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE; 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DOMAIN; 961 1443 10 X TANDEM REPEATS, CYS-RICH.  
 FT REPEAT; 961 1006 1.  
 FT REPEAT; 1007 1056 2.  
 FT REPEAT; 1057 1103 3.  
 FT REPEAT; 1104 1152 4.  
 FT REPEAT; 1153 1204 5.  
 FT REPEAT; 1205 1253 6.  
 FT REPEAT; 1254 1298 7.  
 FT REPEAT; 1299 1345 8.  
 FT REPEAT; 1346 1392 9.  
 FT REPEAT; 1393 1443 10.  
 FT TRANSMEM; 1512 1532 POTENTIAL.  
 FT DOMAIN; 1533 1679 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD; 3 3 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 109 109 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 130 130 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 205 205 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 442 442 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 480 480 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 927 927 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 1060 1060 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 1181 1181 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 1274 1274 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 1277 1277 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD; 1439 1439 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CONFLICT; 152 153 MISSING (IN REF. 1).  
 FT CONFLICT; 177 177 V -> F (IN REF. 1).  
 FT CONFLICT; 213 213 V -> VDQL (IN REF. 1).  
 SQ SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CF6 CRC64;

Query Match 11.68; Score 176.5; DB 1; Length 1679;  
 Best Local Similarity 28.08; Pred.No. 3e-05;  
 Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

QY 3 HRLISWLPFIILNFMFYIGSNASRRRRORRHPNVSGQGGGATCDY-NGCLSCKPR 61  
 DB 1037 HLHVID-LAVCLQCPDGYFENS-----NRRTCPV-----CEPNCASCODHPEYCTSCDHH 1086

QY 62 LPFALERIGMKQGVCLSSCPGSGYGYTRYDPDINKCTYKADCDTCF--NNPCTKCSGF 119  
 DB 1087 LVNHEHK-----CYSACPLDYET---BDNKCACFCHSTCATCNGPTDQDCTTCSSR 1135

QY 120 YLHLGKCLDNCPEGLRANNHMECVSVHCVSEWNPSPCTKGGKTCGFKRGVETEV-R 178  
 DB 1136 YAWQNKLCISCPDGYADKKRLECM-----PCQRCCKTC-----TSNGVCS 1176

QY 179 EIIQHPSAKGNLCPPPTNTRKCTVQRKK-CQKGE 211  
 DB 1177 ECLQNWIT-----LNKRDKCVSGSGGCSSEB 1202

RESULT 4  
 PAC4 HUMAN STANDARD; PRT: 969 AA.  
 AC P29132; Q15099; Q15100; Q9UEG7; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8;  
 AC Q9UEJ9; Q9Y4G9; Q9Y4H0; Q9Y4H1;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)  
 DE (Subtilisin/kexin-like protease PAC4) (Subtilisin-like proprotein  
 DE convertase 4) (SPC4).  
 OS PAC4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I AND PAC4B).  
 RC TISSUE=Hepatoma, and Kidney;  
 RX MEDLINE=92075167; PubMed=1741956;  
 RA Kiefer M.C., Tucker J.B., Joh R., Landeberg K.B., Saltman D.,  
 Barr P.J.;  
 RA "Identification of a second human subtilisin-like protease gene in  
 RT the tes/fps region of chromosome 15";  
 RT DNA Cell Biol. 10:757-769(1991).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I AND PAC4D).  
 RC TISSUE=Placenta;  
 RX MEDLINE=94235049; PubMed=8179631;  
 RA Teuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,  
 Matsuda Y.;  
 RA "Identification of novel cDNAs encoding human kexin-like protease,  
 RT PAC4 isoforms";  
 RT Biochem. Biophys. Res. Commun. 200:943-950(1994).  
 [3]  
 RN ERRATUM.  
 RP MEDLINE=95071480; PubMed=7980617;  
 RA Teuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,  
 Matsuda Y.;  
 RA "Identification of novel cDNAs encoding human kexin-like protease,  
 RT PAC4 isoforms";  
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).  
 [4]  
 RN SEQUENCE FROM N.A. (ISOFORM PAC4A-II).  
 RC TISSUE=Placenta;  
 RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Teuji A.,  
 Matsuda Y.;  
 RA "Identification of a novel PAC4 isoform, PAC4E";  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RN SEQUENCE FROM N.A. (ISOFORMS PAC4B-I AND PAC4B-II).  
 RC TISSUE=Cerebellum;

RX MEDLINE=97335942; PubMed=912737;  
 RA Mori K., Kii S., Teuji A., Nagahama M., Imamaki A., Hayashi K.,  
 Akamatsu T., Nagamune H., Matsuda Y.;  
 RA "A novel human PAC4 isoform, PAC4E is an active processing protease  
 RT containing a hydrophobic cluster at the carboxy terminus";  
 RL J. Biochem. 121:941-948(1997).  
 [6]  
 RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I; PAC4A-II; PAC4C; PAC4D;  
 RP PAC4E-I AND PAC4E-II).  
 RX MEDLINE=96021089; PubMed=9378725;  
 RA Teuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,  
 Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;  
 RA "Genomic organization and alternative splicing of human PAC4 (SPC4),  
 RT kexin-like processing endoprotease";  
 RL J. Biochem. 122:438-452(1997).  
 [7]  
 RN ALTERNATIVE SPLICING (ISOFORM PAC4CS).  
 RP MEDLINE=97064242; PubMed=8906861;  
 RX Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;  
 RA "Functional analysis of human PAC4-A and PAC4-C isoforms:  
 RT identification of a new PAC4-CS isoform";  
 RL FEBS Lett. 396:31-36(1996).  
 [8]  
 RN CHARACTERIZATION.  
 RP MEDLINE=99233559; PubMed=10215603;  
 RX Sucic J.F., Moehring J.M., Inocencio N.M., Luchini J.W.,  
 Moehring T.J.;  
 RA "Endoprotease PAC4 is Ca<sup>2+</sup>-dependent and temperature-sensitive and  
 RT can partly rescue the phenotype of a furin-deficient cell strain";  
 RL Biochem. J. 339:639-647(1999).  
 [9]  
 RN PROCESSING.  
 RP MEDLINE=96408849; PubMed=9738469;  
 RX Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,  
 Teuji A., Matsuda Y.;  
 RA "Biosynthetic processing and quaternary interactions of proprotein  
 RT convertase SPC4 (PAC4)";  
 RL FEBS Lett. 434:155-159(1998).  
 CC -I- FUNCTION: Likely to represent an endoprotease activity within the  
 CC constitutive secretory pathway, with unique restricted  
 CC distribution in both neuroendocrine and non-neuroendocrine tissues  
 CC and capable of cleavage at the RX(K/R)R consensus motif.  
 CC -I- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -I- COFACTOR: Calcium (potential).  
 CC -I- SUBUNIT: The PAC4A-I precursor protein seems to exist in the  
 CC reticulum endoplasmic as both a monomer and a dimer-sized complex  
 CC whereas mature PAC4A-I exists only as a monomer, suggesting that  
 CC propeptide cleavage affects its tertiary or quaternary structure.  
 CC and PAC4CS are not secreted and remain probably in zymogen form  
 CC in endoplasmic reticulum. PAC4B-I and PAC4E-II are retained  
 CC intracellularly probably through a hydrophobic cluster in their C-  
 CC terminus. PAC4B might be secreted.  
 CC -I- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Name=PAC4A-I; Synonyms=PAC4;  
 CC IsoId=P29122-1; Sequence=Displayed;  
 CC Name=PAC4A-II;  
 CC IsoId=P29122-2; Sequence=VSP\_005436;  
 CC Name=PAC4B; Synonyms=PAC4A.1;  
 CC IsoId=P29122-3; Sequence=VSP\_005428, VSP\_005429;  
 CC Note=Probably enzymatically inactive;  
 CC Name=PAC4C;  
 CC IsoId=P29122-4; Sequence=VSP\_005432, VSP\_005433;  
 CC Note=Probably enzymatically inactive;  
 CC Name=PAC4CS;  
 CC IsoId=P29122-5; Sequence=VSP\_005430, VSP\_005431;  
 CC Note=Probably enzymatically inactive;  
 CC Name=PAC4D;  
 CC IsoId=P29122-6; Sequence=VSP\_005427, VSP\_005434, VSP\_005435;  
 CC Note=Probably enzymatically inactive;

CC Name=PACB4E-I; IsoId=P29122-7; Sequence=VSP\_005437;  
 CC Name=PACB4E-II; IsoId=P29122-8; Sequence=VSP\_005436, VSP\_005437;  
 CC TISSUE SPECIFICITY: Each PACB4E isoform exhibits a unique  
 CC restricted distribution. PACB4E-I is expressed in heart, brain,  
 CC placenta, lung, skeletal muscle, kidney, pancreas, but at  
 CC comparatively higher levels in the liver. PACB4E-II is at least  
 CC expressed in placenta. PACB4B was only found in the embryonic  
 CC kidney cell line from which it was isolated. PACB4C and PACB4D are  
 CC expressed in placenta. PACB4E-I is expressed in cerebellum,  
 CC placenta and pituitary. PACB4E-II is at least present in  
 CC cerebellum.  
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
 CC assisting the folding of the zymogen within the endoplasmic  
 CC reticulum. Isoform PACB4D lacks the propeptide domain.  
 CC -!- SIMILARITY: Belongs to peptidase family S8.  
 CC -!- SIMILARITY: Contains 1 homo B/P domain.  
 CC -!- SIMILARITY: Contains 1 PLAC domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; M80482; AAA59998.1; -;  
 CC DR EMBL; AB001914; BAA21620.1; JOINED.  
 CC DR EMBL; AB001898; BAA21620.1; JOINED.  
 CC DR EMBL; AB001900; BAA21620.1; JOINED.  
 CC DR EMBL; AB001901; BAA21620.1; JOINED.  
 CC DR EMBL; AB001902; BAA21620.1; JOINED.  
 CC DR EMBL; AB001903; BAA21620.1; JOINED.  
 CC DR EMBL; AB001904; BAA21620.1; JOINED.  
 CC DR EMBL; AB001905; BAA21620.1; JOINED.  
 CC DR EMBL; AB001914; BAA21621.1; JOINED.  
 CC DR EMBL; AB001898; BAA21621.1; JOINED.  
 CC DR EMBL; AB001900; BAA21621.1; JOINED.  
 CC DR EMBL; AB001901; BAA21621.1; JOINED.  
 CC DR EMBL; AB001902; BAA21621.1; JOINED.  
 CC DR EMBL; AB001903; BAA21621.1; JOINED.  
 CC DR EMBL; AB001904; BAA21621.1; JOINED.  
 CC DR EMBL; AB001905; BAA21621.1; JOINED.  
 CC DR EMBL; AB001906; BAA21621.1; JOINED.  
 CC DR EMBL; AB001907; BAA21621.1; JOINED.  
 CC DR EMBL; AB001908; BAA21621.1; JOINED.  
 CC DR EMBL; AB001909; BAA21621.1; JOINED.  
 CC DR EMBL; AB001914; BAA21622.1; JOINED.  
 CC DR EMBL; AB001901; BAA21622.1; JOINED.  
 CC DR EMBL; AB001902; BAA21622.1; JOINED.  
 CC DR EMBL; AB001903; BAA21622.1; JOINED.  
 CC DR EMBL; AB001904; BAA21622.1; JOINED.  
 CC DR EMBL; AB001905; BAA21622.1; JOINED.  
 CC DR EMBL; AB001906; BAA21622.1; JOINED.  
 CC DR EMBL; AB001907; BAA21622.1; JOINED.  
 CC DR EMBL; AB001908; BAA21622.1; JOINED.  
 CC DR EMBL; AB001914; BAA21623.1; JOINED.  
 CC DR EMBL; AB001898; BAA21623.1; JOINED.  
 CC DR EMBL; AB001900; BAA21623.1; JOINED.  
 CC DR EMBL; AB001901; BAA21623.1; JOINED.  
 CC DR EMBL; AB001902; BAA21623.1; JOINED.  
 CC DR EMBL; AB001903; BAA21623.1; JOINED.  
 CC DR EMBL; AB001904; BAA21623.1; JOINED.  
 CC DR EMBL; AB001905; BAA21623.1; JOINED.  
 CC DR EMBL; AB001906; BAA21623.1; JOINED.  
 CC DR EMBL; AB001907; BAA21623.1; JOINED.  
 CC DR EMBL; AB001908; BAA21623.1; JOINED.  
 CC DR EMBL; AB001914; BAA21624.1; JOINED.  
 CC DR EMBL; AB001898; BAA21624.1; JOINED.  
 CC DR EMBL; AB001900; BAA21624.1; JOINED.

DR EMBL; AB001901; BAA21624.1; JOINED.  
 DR EMBL; AB001902; BAA21624.1; JOINED.  
 DR EMBL; AB001903; BAA21624.1; JOINED.  
 DR EMBL; AB001904; BAA21624.1; JOINED.  
 Query Match 11.4%; Score 173.5; DB 1; Length 969;  
 Best Local Similarity 28.8%; Pred. No. 2.9e-05;  
 Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;  
 QY 19 YTGQNASRGRQRMRHMFNVSGCQGCGATCSP--YNGCLSCXPRLPFALRIGRMKQIGV 76  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 737 YFGDTAARRCR-----CHGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779  
 QY 77 CLSSCPGSGYGTYPDINKCTKADCDTCFNK-NCTKCKSGFYHLHGLKLDNCPGLE 135  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 780 CVTLCPAGFYAD--SQKCLKCHPSCKVCVDPKCTVCKGFSLGSGCIPDCPEGY 837  
 QY 136 ANNHTMECVSIVHCEVSEWNPSPCTKGGKTCG 168  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 838 FDSLRGCECH-----TCG 853  
 RESULT 5  
 PCK5 RAT  
 ID\_PCK5 RAT STANDARD; PRT; 1877 AA.  
 AC P41413; Q62914;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Protease convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
 DE (Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
 DE (zPCS) (PC6) (Fragment).  
 DE PCSK5.  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.  
 RC TISSUE=Adrenal gland;  
 RX MEDLINE=93342056; PubMed=8341687;  
 RA Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
 RT "CDNA structure of the mouse and rat subtilisin/kexin-like PCS; a  
 RT candidate proprotein convertase expressed in endocrine and  
 RT nonendocrine cells";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
 RN [2]  
 RP SEQUENCE OF 1676-1877 FROM N.A. (ISOFORM LONG).  
 RC TISSUE=Adrenal gland;  
 RA De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=20214819; PubMed=10749928;  
 RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;  
 RT "The PC68 cytoplasmic domain contains two acidic clusters that direct  
 RT sorting to distinct trans-Golgi network/endosomal compartments.";  
 RL Mol. Biol. Cell 11:1257-1273(2000).  
 RN [4]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=97166043; PubMed=9013936;  
 RA Zheng M., Seidah N.G., Pintar J.E.;  
 RT "The developmental expression in the rat CNS and peripheral tissues of  
 RT proteases PCS and PACE4 mRNAs: comparison with other proprotein  
 RT processing enzymes.";  
 RL Dev. Biol. 181:268-283(1997).  
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity  
 CC within the constitutive and regulated secretory pathway. Capable  
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible  
 CC for the maturation of gastrointestinal peptides. May be involved  
 CC in the cellular proliferation of adrenal cortex via the activation  
 CC of growth factors.  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their





RA Reudelhuber T.L.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity  
 CC within the constitutive and regulated secretory pathway. Capable  
 CC of cleavage at the RX(K/R)R consensus motif.  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=1;  
 CC Comment=2 isoforms may be produced;  
 CC Name=1;  
 CC IsoId=Q92824-1; Sequence=Displayed;  
 CC TISSUE SPECIFICITY: Expressed in T-lymphocytes.  
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
 CC assisting the folding of the zymogen within the endoplasmic  
 CC reticulum.  
 CC -!- SIMILARITY: Belongs to peptidase family S8.  
 CC -!- SIMILARITY: Contains 1 homo B/P domain.  
 CC -!- SIMILARITY: Contains 1 PLAC domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U56387; AAC50643.2; --  
 CC EMBL; U49114; AAA91807.1; --  
 CC PIR; G02428; G02428.  
 CC HSP; Q99405; 1MPT.  
 CC MEROPS; S08.076; --  
 CC Genew; HGNC:8747; PCSK5.  
 CC MIM; 600488; --  
 CC GO; GO:0005615; C:extracellular space; TAS.  
 CC GO; GO:0004252; F:serine-type endopeptidase activity; TAS.  
 CC GO; GO:007267; P:cell-cell signaling; TAS.  
 CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
 CC InterPro; IPR006212; Furin repeat.  
 CC InterPro; IPR000209; Peptidase S8.  
 CC InterPro; IPR002884; Peptidase\_S8B.  
 CC InterPro; IPR009020; Protease\_inhib.  
 CC Pfam; PF01483; P:protease; 1.  
 CC Pfam; PF00082; Peptidase S8; 1.  
 CC PRINTS; PR00723; SUBTILISIN.  
 CC ProDom; PD000717; P domain; 1.  
 CC SMART; SM00261; FU\_5.  
 CC PROSITE; PS00900; PLAC; 1.  
 CC PROSITE; PS00136; SUBTILASE ASP; 1.  
 CC PROSITE; PS00137; SUBTILASE HIS; 1.  
 CC PROSITE; PS00138; SUBTILASE\_SER; 1.  
 CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing.  
 FT SIGNAL 1 32  
 FT PROPEP 33 114 BY SIMILARITY.  
 FT CHAIN 115 913 PROPEPTIDE CONVERTASE SUBTILISIN/KEXIN  
 FT TYPE 5.  
 FT DOMAIN 115 454 CATALYTIC.  
 FT DOMAIN 462 600 HOMO B.  
 FT DOMAIN 636 868 CYS-RICH MOTIF (CRM) REGION.  
 FT DOMAIN 869 913 PLAC.  
 FT SITE 114 115 CLEAVAGE (AUTO-) (BY SIMILARITY).  
 FT SITE 519 521 CELL ATTACHMENT SITE (POTENTIAL).  
 FT ACT\_SITE 171 171 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 386 386 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 225 225 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 852 852 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 118 118 S -> F (IN REF. 3).  
 FT CONFLICT 121 121 V -> A (IN REF. 3).  
 FT CONFLICT 511 511 R -> A (IN REF. 3).  
 FT CONFLICT 601 601 R -> Q (IN REF. 3).  
 SQ SEQUENCE 913 AA; 101775 MW; 21389264CAD7546C CRC64;  
 Query Match 10.7%; Score 162; DB 1; Length 913;  
 Best Local Similarity 22.3%; Pred. No. 0.00018;  
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;  
 QY 21 GSONASRGRRQRMRHNVSGGGGCGATCSDYNG--CLSCPKRLFPALERIGMKQIGVCL 78  
 DB 677 GHYHAK-KRCRKAPN-----CSCPGSHGDDQMSCKYGYFL-----NBEITNSCV 721  
 QY 79 SSCPSG-YGTTRYPDINKCTCKNACDCTCFNKNFCTKCKSGPYLHLGLCLDNCPEGLEAN 137  
 DB 722 THCPDGSYQTKK---NLCKSCSECKTCTCFHCTCTCDGLSLQGRSCSVSCDGRYFN 778  
 QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPCTKK 163  
 DB 779 GDCQCPCHRFCAFCAGAGAGGCGINCCTGYFMEDGRCVQSCSISYFPHSSSENGYKCKKC 838  
 QY 164 GKTC-----GPKRGTE-----TRVEIIGHPSAKGNLCPTTETKCTV 202  
 DB 839 DRISLCTNGPGFNCCTSCPSGYLLDGMQMGAIKDAYEESWARGGFCMLVKXNLC-- 896  
 QY 203 QRKCKQK 209  
 DB 897 QRKVLQK 903  
 RESULT 7  
 ID WIF1 HUMAN STANDARD; PRT; 379 AA.  
 AC Q9Y5W5; Q8WVG4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Wnt inhibitory factor 1 precursor (WIF-1).  
 GN WIF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Heich J.-C., Kodjabachian L., Rebber M.L., Rattner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., Dawid I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 RT activities.";  
 RL Nature 398:431-436(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,  
 RA Altshul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., MCSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.B.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
 CC -!- FUNCTION: Binds to Wnt proteins and inhibits their activities. May  
 be involved in mesoderm segmentation.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Contains 5 EGF-like domains.  
 CC -!- SIMILARITY: Contains 1 WIF domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF122922; AAD25402.1; --  
 DR EMBL; BC018037; AAH18037.1; --  
 DR PIR; A59180; A59180.  
 DR HSP; P00743; LCCP.  
 DR Genew; HGNC:18081; WIF1.  
 DR MIM; 605186; --  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR003306; WIF.  
 DR Pfam; PF00008; EGF; 5.  
 DR Pfam; PF02019; WIF; 1.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR SMART; SM00181; EGF; 5.  
 DR SMART; SM00469; WIF; 1.  
 DR PROSITE; PS00022; EGF\_1; 5.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00026; EGF\_3; 5.  
 DR PROSITE; PS00814; WIF; 1.  
 KW Repeat; EGF-like domain; Signal; Developmental protein;  
 KW Wnt signaling pathway.  
 FT SIGNAL 1 28  
 FT CHAIN 29 379  
 FT DOMAIN 38 176  
 FT DOMAIN 177 208  
 FT DOMAIN 209 240  
 FT DOMAIN 241 272  
 FT DOMAIN 273 304  
 FT DOMAIN 305 336  
 FT DISULFID 177 186  
 FT DISULFID 182 192  
 FT DISULFID 198 200  
 FT DISULFID 209 218  
 FT DISULFID 214 224  
 FT DISULFID 230 232  
 FT DISULFID 241 250  
 FT DISULFID 246 256  
 FT DISULFID 262 264  
 FT DISULFID 273 282  
 FT DISULFID 278 288  
 FT DISULFID 294 296  
 FT DISULFID 305 314  
 FT DISULFID 310 320  
 FT DISULFID 326 328  
 FT CARBOHYD 88 88  
 FT CARBOHYD 245 245  
 FT CONFLICT 166 166  
 FT CONFLICT 178 178  
 SQ SEQUENCE 379 AA; 41528 MW; 32EC54D60529EP96 CRC64;  
 Query Match 10.5%; Score 158.5; DB 1; Length 379;  
 Best Local Similarity 24.9%; Pred. No. 0.00014;  
 Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

QY 42 CCGCAT--CSDYNGC-----LSCPRLPFALEIRGNKQIGVCLSS---C 81  
 DB 182 CPGCRNGFCNERRICECPDGFPHGPHCEKALCTPONG-----GLCVTPGFCIC 232  
 QY 82 PSYGYTRYDINKTKC-KADCD--TCFNKPNFTCKSGFYHLGKCLDNCPEGLRANH 139  
 DB 233 PPGFVG-----VNCDKANCSITCFNGGTC-----FY--PGKCI--CPGLEGR-- 271  
 QY 140 TMECVSIVHCEVSEWNPSPCTKKGKCTGPKRGTTETRVREIIQHPSAKGNLCPP----- 193  
 DB 272 -----QCHSKCP--QPCRNKGKCTG-----KSKCKSGYGGDLCKSPVCEPGCGANG 318  
 QY 194 -TNETRTKTVQKKCKQKGERGKGRERK-----KFNKGSKEAIPDSKLSKSKBP 246  
 DB 319 TCHEPNK-----CQEGHGRHCNKRYEASLIHALRPAGALRQHTPSLKAERDRDP 373  
 QY 247 E 247  
 DB 374 E 374

RESULT 8  
 WIF1\_XENLA  
 ID WIF1\_XENLA STANDARD; PRT; 374 AA.  
 AC Q9W6F8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DB Wnt inhibitory factor 1 precursor (WIF-1).  
 GN WIF1.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Hsieh J.-C., Kodjabachian L., Rebbert M.L., Rattner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.:  
 RA "A new secreted protein that binds to Wnt proteins and inhibits their  
 activities.";  
 RT Nature 398:431-436(1999).  
 RL Nature 398:431-436(1999).  
 CC -!- FUNCTION: Binds to Wnt proteins and inhibits their activities. May  
 be involved in mesoderm segmentation.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: During somitogenesis, expressed predominantly  
 in unsegmented paraxial presomitic mesoderm and, to a much lesser  
 extent, in newly segmented somites.  
 CC -!- DEVELOPMENTAL STAGE: First expressed at neurula stages.  
 CC -!- SIMILARITY: Contains 5 EGF-like domains.  
 CC -!- SIMILARITY: Contains 1 WIF domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF122924; AAD25404.1; --  
 DR HSP; P00740; IEDM.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR003306; WIF.  
 DR Pfam; PF00008; EGF; 5.  
 DR Pfam; PF02019; WIF; 1.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR SMART; SM00181; EGF; 4.  
 DR SMART; SM00469; WIF; 1.  
 DR PROSITE; PS00022; EGF\_1; 5.

```

DR PROSITE; PS01186; EGF 2; 5.
DR PROSITE; PS00226; EGF_3; 4.
DR PROSITE; PS0814; WIF; 1.
KW Repeat; EGF-like domain; Signal; Developmental protein;
Wnt signaling pathway.
FT SIGNAL 1 28
FT CHAIN 29 374
FT DOMAIN 33 171
FT DOMAIN 172 203
FT DOMAIN 204 235
FT DOMAIN 236 267
FT DOMAIN 268 299
FT DOMAIN 300 331
FT DISULFID 172 181
FT DISULFID 177 187
FT DISULFID 193 195
FT DISULFID 204 213
FT DISULFID 209 219
FT DISULFID 225 227
FT DISULFID 236 245
FT DISULFID 241 251
FT DISULFID 257 259
FT DISULFID 268 277
FT DISULFID 273 283
FT DISULFID 289 291
FT DISULFID 300 309
FT DISULFID 305 315
FT DISULFID 321 323
FT CARBOHYD 83 83
FT CARBOHYD 240 240
SQ SEQUENCE 374 AA; 41071 MW; E26F973B0F00ACF8 CRC64;

Query Match
Best Local Similarity 10.4%; Score 157.5; DB 1; Length 374;
Matches 57; Conservative 22; Mismatches 85; Indels 71; Gaps 12;

QY 42 CQGGCAT---CSDYNGC-----LSCKPRLFPALERIGMKQIGVC-----LSSC 81
DB 177 CTGCGRNGFGFCDNRHVCCECDPGFGPHCEKALCMPCWNG-----GLCTPGLCIC 227
QY 82 PSQYGYTRYPDINKTKCKADCTCFKNKFNCTKSGFYHLGLKCLDNCEGLEANNHTM 141
DB 228 PPGYGINCDKVNCTTHC-----YPGKCI--CPSGYEGE----- 266
QY 142 ECYSIVHCEVSEWNPSPCKTKGCTGPKGTGTRVRELIHPSAKGNLCPTNETKCT 201
DB 267 -----QCETSKQ--QPCRNGGKCSG---KNKCKSGKYQGDLCNKPVCPCSGAHGTC 315
QY 202 VQRKCKQGRGKGRERKPK-----KPNKGSKEAIPDSKLSSESKPE 247
DB 316 IEPNKCQCKE-GWNGRYCNKKYGSNLNMLNARPTGSRNRQHTPSPKRTEDQALPE 369

RESULT 9
PAC4 RAT STANDARD; PRT; 937 AA.
AC Q63415;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Paired basic amino acid cleaving enzyme 4 precursor (BC 3.4.21.-)
DE (Subtilisin/kexin-like protease PACB4) (Subtilisin-like propeptide
DE convertase 4) (SPC4).
GN PACB4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Pituitary;
RX MEDLINE=94349873; PubMed=8070361;
RA Johnson R.C., Darlington D.N., Hand T.A., Bloomquist B.T., Mains R.E.;

```

```

RT "PACB4: a subtilisin-like endoprotease prevalent in the anterior
RT pituitary and regulated by thyroid status.";
RL Endocrinology 135:1178-1185(1994).
CC -!- FUNCTION: Likely to represent an endoprotease activity within the
CC constitutive secretory pathway, with unique restricted
CC distribution in both neuroendocrine and non-neuroendocrine tissues
CC and capable of cleavage at the RX(K/R)R consensus motif.
CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys.
CC -!- COFACTOR: Calcium (Potential).
CC -!- TISSUE SPECIFICITY: High expression in the anterior pituitary and
CC in several brain regions, the atrium, and the ventricle.
CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
CC assisting the folding of the zymogen within the endoplasmic
CC reticulum.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -!- SIMILARITY: Contains 1 homo B/P domain.
CC -!- SIMILARITY: Contains 1 PLAC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L31894; AAA61987.1; -.
DR F1R; I53282; I53282.
DR HSP; Q99405; LMPT.
DR MEROPS; S08.075; -.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR InterPro; IPR009020; Protease_inh.
DR Pfam; PF01483; P_protein; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00261; FU; 5.
DR PROSITE; PS0900; PLAC; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal; Calcium;
KW Cleavage on pair of basic residues; Repeat.
FT SIGNAL 1 45
FT PROPEP 46 132
FT CHAIN 133 937
FT DOMAIN 133 454
FT DOMAIN 477 615
FT DOMAIN 680 898
FT DOMAIN 899 937
FT SITE 132 133
FT SITE 534 536
FT ACT_SITE 186 186
FT ACT_SITE 227 227
FT ACT_SITE 401 401
FT CARBOHYD 240 240
FT CARBOHYD 882 882
FT CARBOHYD 900 900
FT SEQUENCE 937 AA; 104053 MW; P3865557C33705C8 CRC64;

Query Match
Best Local Similarity 10.3%; Score 156.5; DB 1; Length 937;
Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11;

QY 19 YIGSNASRGRRQRMRHFNPNVSGCGCATCSDYN--GCLSKPRLFPALERIGMKQIGV 76
DB 705 YFGDTAARRCR-----CHKGCTCTGKSPQCLSCR-RGFI-----HQETNT 747

```

QY 77 CLSCPSGYTRYDINKCTCKKADCTCFNK-NFCTCKSGFYHLGKCLDNCPEGLE 135  
 Db 748 CVTLCPAGLYADSSQL--CLRCHPSCKVCDBPEKSTVCKEGFSLARGSCIPDCBPGTY 805  
 QY 136 ANNHTMEC-----VSIVHCEVS-----EWNPSVCTKKGKTCGFKRGTETR 176  
 Db 806 FDSSELIRCGECHTCTCVGSPREECCHCAKSFHQDMKCVPAQGE-----GP----- 853  
 QY 177 VREIIOHPSAKGNLCPTNETRKTCTVQKCKQGER 212  
 Db 854 -----YPEMPGL--PHKVCRRCDNCLSCGSSR 881

RESULT 10  
 FSPO XENLA STANDARD; PRT; 803 AA.  
 ID FSPO XENLA AC P3547;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE F-spondin precursor.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93376785; PubMed=8367492;  
 RA Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;  
 RT "Ectopic neural expression of a floor plate marker in frog embryos  
 injected with the midline transcription factor Pintallavis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).  
 CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory  
 CC neuron cells and the outgrowth of neurites in vitro. May  
 CC contribute to the growth and guidance of axons in both the spinal  
 CC cord and the PNS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.  
 CC -!- SIMILARITY: Contains 6 TSP type-1 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L09123; AAA19105.1; -.  
 DR PIR; A47723; A47723.  
 DR InterPro; IPR002861; Reeler.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF02014; Reeler; 1.  
 DR Pfam; PF00090; tsp.1; 6.  
 DR SMART; SM00209; TSP1; 6.  
 DR PROSITE; PS50092; TSP1; 6.  
 KW Glycoprotein; Signal; Repeat; Cell adhesion.  
 FT SIGNAL 1 23  
 FT CHAIN 24 803  
 FT DOMAIN 436 489  
 FT DOMAIN 495 549  
 FT DOMAIN 552 605  
 FT DOMAIN 608 662  
 FT DOMAIN 664 717  
 FT DOMAIN 750 802  
 FT CARBOHYD 210 210  
 FT CARBOHYD 677 677  
 SQ SEQUENCE 803 AA; 90702 MW; D3A54E329548AED9 CRC64;

Query Match 10.3%; Score 156; DB 1; Length 803;  
 Best Local Similarity 24.8%; Pred. No. 0.00044;

Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;  
 QY 38 VSQGGCGGATCSDYNGCLSKPRLPALRIGMKQIGVCLSSCPSS--GYGTRYPDINK 95  
 Db 544 VNEBESPSSCIVTEWAEWBECS-----ATCRMGKRRHMKMTPADGSMCKADTTVEVK 598  
 QY 96 C--TKK-----ADCTCFNKNFCTCKSGFYHLGKCLDNCPEGLEANN--- 138  
 Db 599 CMPECHTIPCVLSPWSEWSDCVTCGKTRTRQR---MLKSPSELGDCNELELKVQEK 655  
 QY 139 -HMECVSVIHCEVSENNPSPCTKCKTCTGFKGTETRVREIIQHPSAKGNLCPTNET 197  
 Db 656 CMLPECP--ISCLTENSYSWSEC---NKSG--KGMIRTEMTWEPQFGAVCPETVQR 708  
 QY 198 RKCTVQKCKQGERGKCKGRKRKPKNKGESKEAIPDSKLSKSKSIPRENNKQ 253  
 Db 709 KKCL--RKCK-----SSGNERRHLK-----DAREKRSEKIKEDSDGQ 747

RESULT 11  
 NTC2 MOUSE STANDARD; PRT; 2470 AA.  
 ID NTC2 MOUSE AC Q35516; Q60941;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch  
 DE B).  
 DE NOTCH2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Rodentia; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Thymus;  
 RA Hamada Y., Higuchi M., Tsujimoto Y.;  
 RT "Complete amino acid sequence and multifunctional transcripts encoded by a  
 RT single copy of mouse Notch2 gene.";  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 316-1518 FROM N.A.  
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
 RX MEDLINE=93178563; PubMed=8440332;  
 RA Lardelli M., Lendahl U.;  
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a  
 RT wide variety of tissues.";  
 RL Exp. Cell Res. 204:364-372(1993).  
 RN [3]  
 RP SEQUENCE OF 1765-2153 FROM N.A.  
 RX MEDLINE=97075110; PubMed=8917536;  
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,  
 RA Martin D.I.;  
 RT "Inhibition of granulocytic differentiation by mNotch1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=99396706; PubMed=10393120;  
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,  
 RA Tsujimoto Y.;  
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
 RT embryonic lethality.";  
 RL Development 126:3415-3424(1999).  
 RN [5]  
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.  
 RX MEDLINE=95333893; PubMed=7609614;  
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;  
 RT "Differential expression of Notch1 and Notch2 in developing and adult  
 RT mouse brain.";  
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).  
 RN [6]  
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
 RX MEDLINE=21523956; PubMed=11518718;

RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (NL-4) undergo presenilin-dependent  
RL proteolysis";  
RN J. Biol. Chem. 276:40268-40273 (2001).  
[7]  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MBT-1699.  
RX MEDLINE=21374376; PubMed=1145941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RL among mammalian Notch family members";  
RC Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).  
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (by similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation.  
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(SC) which are probably linked by disulfide  
CC bonds.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=O35516-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=O35516-2; Sequence=VSP\_001405;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
CC neuroepithelial, somites, optic vesicles and branchial arches, but  
CC not heart.  
CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
CC the postnatal ependymal cells, and the choroid plexus throughout  
CC embryonic and postnatal development.  
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(SC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (TACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC -!- PTM: Phosphorylated.  
CC -!- SIMILARITY: Belongs to the NOTCH family.  
CC -!- SIMILARITY: Contains 35 EGF-like domains.  
CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.  
CC -!- SIMILARITY: Contains 6 ANK repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: D32210; BAA22094.1; -;  
CC EMBL: X68279; CAA48340.1; -;  
CC EMBL: U11881; AAC52924.1; -;  
CC PIR: A49175; A49175.  
CC HSP: P16109; IFSB.  
CC MGD: MGI:97364; Notch2.  
CC GO: GO:0005887; C: integral to plasma membrane; IC.  
CC GO: GO:0005515; P: protein binding; IPT.  
CC GO: GO:0002011; P: morphogenesis of an epithelial sheet; IMP.  
CC GO: GO:0007219; P: N signaling pathway; IC.

DR InterPro: IPR002110; ANK.  
DR InterPro: IPR000152; Asx\_hydroxyl\_S.  
DR InterPro: IPR000742; EGF\_2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_II.  
DR InterPro: IPR006209; EGF\_like.  
DR InterPro: IPR002049; Laminin\_EGF.  
DR InterPro: IPR008297; Notch.  
DR InterPro: IPR000800; Notch\_dom.  
DR Pfam: PF00023; ank; 6.  
DR Pfam: PF00008; EGF; 34.  
DR Pfam: PF00066; notch; 2.  
DR PIRSF: PIRSF002279; Notch; 1.  
DR PRINTS: PRO0010; EGFBLD.  
DR PRINTS: PRO0011; EGF\_LAMININ.  
DR PRINTS: PRO1452; NOTCH.  
DR SMART: SM00248; ANK; 6.  
DR SMART: SM00179; EGF\_Ca; 23.  
DR SMART: SM00004; NL; 3.  
DR PROSITE: PS50297; ANK\_REPEAT; 1.  
DR PROSITE: PS00088; ANK\_REPEAT; 4.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
DR PROSITE: PS00022; EGF\_1; 33.  
DR PROSITE: PS01186; EGF\_2; 27.  
DR PROSITE: PS50026; EGF\_3; 35.  
DR PROSITE: PS01187; EGF\_Ca; 22.  
DR Receptor; transcription regulation; Activator; Differentiation;  
DR Developmental protein; repeat; ANK repeat; EGF-like domain;  
DR Transmembrane; Glycoprotein; Signal; Phosphorylation;  
DR Alternative splicing.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.  
FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.  
FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1678 1698 POTENTIAL.  
FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 26 63 EGF-LIKE 1.  
FT DOMAIN 64 102 EGF-LIKE 2.  
FT DOMAIN 105 143 EGF-LIKE 3.  
FT DOMAIN 144 180 EGF-LIKE 4.  
FT DOMAIN 182 219 EGF-LIKE 5.  
FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).  
FT DOMAIN 258 294 EGF-LIKE 7.  
FT DOMAIN 296 334 EGF-LIKE 8.  
FT DOMAIN 336 372 EGF-LIKE 9.  
FT DOMAIN 373 411 EGF-LIKE 10.  
FT DOMAIN 413 452 EGF-LIKE 11.  
FT DOMAIN 454 490 EGF-LIKE 12.  
FT DOMAIN 492 528 EGF-LIKE 13.  
FT DOMAIN 530 566 EGF-LIKE 14.  
FT DOMAIN 568 603 EGF-LIKE 15.  
FT DOMAIN 605 641 EGF-LIKE 16.  
FT DOMAIN 643 678 EGF-LIKE 17.  
FT DOMAIN 680 716 EGF-LIKE 18.  
FT DOMAIN 718 753 EGF-LIKE 19.  
FT DOMAIN 755 791 EGF-LIKE 20.  
FT DOMAIN 793 829 EGF-LIKE 21.  
FT DOMAIN 831 869 EGF-LIKE 22.  
FT DOMAIN 871 907 EGF-LIKE 23.  
FT DOMAIN 909 945 EGF-LIKE 24.  
FT DOMAIN 947 983 EGF-LIKE 25.  
FT DOMAIN 985 1021 EGF-LIKE 26.  
FT DOMAIN 1023 1059 EGF-LIKE 27.  
FT DOMAIN 1061 1097 EGF-LIKE 28.  
FT DOMAIN 1099 1145 EGF-LIKE 29.  
FT DOMAIN 1147 1183 EGF-LIKE 30.  
FT DOMAIN 1185 1221 EGF-LIKE 31.  
FT DOMAIN 1223 1260 EGF-LIKE 32.  
FT DOMAIN 1262 1300 EGF-LIKE 33.  
FT DOMAIN 1302 1345 EGF-LIKE 34.  
FT DOMAIN 1372 1410 EGF-LIKE 35.  
FT REPEAT 1418 1454 LIN/NOTCH 1.

```

FT REPEAT 1501 1533 LIN/NOTCH 2.
FT REPEAT 1825 1869 ANK 1.

Query Match 9.9%; Score 150.5; DB 1; Length 2470;
Best Local Similarity 24.3%; Pred. No. 0.0031;
Matches 53; Conservative 23; Mismatches 53; Indels 89; Gaps 12;

QY 31 ORWHFNVSGCGGNCATCSYNGCLSCPKPLFPALBRIGMKQIGVCLSSCPGSGYCTRY 90
DB 945 QTMNECLSPCKNG-GTCSYNSYTC-----TCPAGFHGVHC 982

QY 91 P-DINKTKCKADCDTCFNKFC-----TKCKSGPYLHLGK 125
DB 983 ENNIDECYE-----SSCFNGGTVDGINSFSLCPVGTGPFCLHDINECSSNPLNAGT 1037

QY 126 CLDN-----CPGLEANNHTMECVSIHV-CEVSENVNPSPTCKKTKCTGFKRGTTTRV 177
DB 1038 CVDGLGYTCICPLGYTGKN---CQTLNVLCS-----RSPCKNGK-TC----- 1076

QY 178 REIIQHPKAGKNCPLPTNETKCTVQKKC-----QKG 210
DB 1077 ---VQKARPHCLCPGPGWGCAYCDVLNVSCRAALQKG 1111

RESULT 12
IGIR RAT
ID IGIR RAT STANDARD; PRT; 1370 AA.
AC P24062;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin-like growth factor I receptor precursor (BC 2.7.1.112).
GN IGIR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95277910; PubMed=7758167;
RA Du J., Delafontaine P.;
RT "Inhibition of vascular smooth muscle cell growth through antisense
transcription of a rat insulin-like growth factor I receptor cDNA.";
RL Circ. Res. 76:963-972(1995).
RN [2]
RP SEQUENCE OF 1-364 FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=50017496; PubMed=2477843;
RA Warner H., Woloschak M., Adamo M., Shen-Orr Z., Roberts C.T. Jr.,
RA Lerioth D.;
RT "Developmental regulation of the rat insulin-like growth factor I
receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7451-7455(1989).
RN [3]
RP SEQUENCE OF 913-1017 FROM N.A.
RX MEDLINE=92412145; PubMed=1530648;
RA Kurachi H., Jojo K., Ohta M., Kawasaki T., Itoh N.;
RT "A new member of the insulin receptor family, insulin
receptor-related receptor, is expressed preferentially in the
kidney.";
RL Biochem. Biophys. Res. Commun. 187:934-939(1992).
CC -1- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)
WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A
TYROSINE-PROTEIN KINASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide
bonds. The alpha chains contribute to the formation of the ligand-
binding domain, while the beta chain carries the kinase domain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin
receptor subfamily.

```

```

CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
CC
CC EMBL; L29232; AAA41392.1; -
CC EMBL; M27293; AAA41384.1; -
CC HSP; P06213; IIRK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR006211; Furin-like.
CC InterPro; IPR006212; Furin repeat.
CC InterPro; IPR009030; Grow_fac_recep.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002011; RecepttyrkinaseII.
CC InterPro; IPR001245; Tyr_kinase.
CC InterPro; IPR008266; Tyr_kinase_AS.
CC Pfam; PF00041; fn3; 2.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recept_L_domain; 2.
CC PRINTS; PR00109; TYRKINASE.
CC PRODOM; PD000001; Prot_kinase; 1.
CC SMART; SM00060; FN3; 3.
CC SMART; SM00261; FU; 1.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 741
FT FT 742 1370
FT CHAIN 742 1370
FT DOMAIN 742 936
FT TRANSHEM 937 960
FT DOMAIN 961 1370
FT DOMAIN 608 829
FT DOMAIN 830 929
FT DOMAIN 1000 1275
FT NP_BIND 1006 1014
FT BINDING 1034 1034
FT ACT_SITE 1136 1136
FT DISULFID 215 224
FT DISULFID 219 230
FT DISULFID 231 239
FT DISULFID 235 248
FT DISULFID 251 260
FT DISULFID 264 276
FT DISULFID 282 303
FT DISULFID 307 321
FT DISULFID 324 328
FT CARBOHYD 51 51
FT CARBOHYD 102 102
FT CARBOHYD 135 135
FT CARBOHYD 245 245
FT CARBOHYD 314 314
FT CARBOHYD 418 418
FT CARBOHYD 439 439
FT CARBOHYD 535 535
FT CARBOHYD 608 608
FT CARBOHYD 623 623
FT CARBOHYD 641 641
FT CARBOHYD 748 748

```

```

-1- SIMILARITY: Contains 2 fibronectin type III domains.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
EMBL; L29232; AAA41392.1; -
EMBL; M27293; AAA41384.1; -
HSP; P06213; IIRK.
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR008957; FN_III-like.
InterPro; IPR003961; FN_III.
InterPro; IPR006211; Furin-like.
InterPro; IPR006212; Furin repeat.
InterPro; IPR009030; Grow_fac_recep.
InterPro; IPR000719; Prot_kinase.
InterPro; IPR002011; RecepttyrkinaseII.
InterPro; IPR001245; Tyr_kinase.
InterPro; IPR008266; Tyr_kinase_AS.
Pfam; PF00041; fn3; 2.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; pkinase; 1.
Pfam; PF01030; Recept_L_domain; 2.
PRINTS; PR00109; TYRKINASE.
PRODOM; PD000001; Prot_kinase; 1.
SMART; SM00060; FN3; 3.
SMART; SM00261; FU; 1.
SMART; SM00219; Tyrc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 741
FT FT 742 1370
FT CHAIN 742 1370
FT DOMAIN 742 936
FT TRANSHEM 937 960
FT DOMAIN 961 1370
FT DOMAIN 608 829
FT DOMAIN 830 929
FT DOMAIN 1000 1275
FT NP_BIND 1006 1014
FT BINDING 1034 1034
FT ACT_SITE 1136 1136
FT DISULFID 215 224
FT DISULFID 219 230
FT DISULFID 231 239
FT DISULFID 235 248
FT DISULFID 251 260
FT DISULFID 264 276
FT DISULFID 282 303
FT DISULFID 307 321
FT DISULFID 324 328
FT CARBOHYD 51 51
FT CARBOHYD 102 102
FT CARBOHYD 135 135
FT CARBOHYD 245 245
FT CARBOHYD 314 314
FT CARBOHYD 418 418
FT CARBOHYD 439 439
FT CARBOHYD 535 535
FT CARBOHYD 608 608
FT CARBOHYD 623 623
FT CARBOHYD 641 641
FT CARBOHYD 748 748

```



```

FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT MOD RES 1166 1166 PHOSPHORYLATION (AUTO-) (BY SIMILARITY) .
FT CONFLICT 985 986 AD -> PY (IN REF. 3) .
SQ SEQUENCE 1370 AA; 155395 MW; A5946897A41CB145 CRC64;

Query Match
Best Local Similarity 26.7%; Pred. No. 0.0023;
Matches 47; Conservative 17; Mismatches 46; Indels 66; Gaps 11;

QY 35 HPNVSGCGGCGATCSDYNGCLSCCKPRLPALRRIGMKQIGVCLSSCPGYY---GTRYP 91
Db 232 HPE---CLGSCHTPDNTTCAVRHYV-----KGVCVPACPPGTYRFGWR-- 275
QY 92 DINKTKKACDCTCNKPNFTKC-----KSGFYHLGKCLDNCPGLEANN-HTMEC 143
Db 276 -----CVRDRFCANIPNAESSDSDFVHDGECMQECPSGFIRNSTQSMYC 321
QY 144 VSIHVCEVSEWNPSPCTKGGTKCGPKRGCTETRVREIIQHPSA-----KGNL 190
Db 322 IP---CE-----GPCP---RVCG---DEKKTITDSVTSQMLQGCTILKGNL 361

RESULT 13
NOTC_XENLA
ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein) .
GN XOTCH.
OS Xenopus laevis (African clawed frog) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.",
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33874; AAB02039.1; -
CC HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_s.
DR InterPro; IPR000742; EGF_2
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002049; Laminin_EGF.

```

```

DR InterPro; IPR008297; Notch.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS02937; ANK_RBP_REGION; 1.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS00101; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS00026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
PT SIGNAL 1 19 POTENTIAL.
PT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
PT TRANSMEM 20 1728 EXTRACELLULAR (POTENTIAL) .
PT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL) .
PT DOMAIN 1 20 57
PT DOMAIN 58 99 EGF-LIKE 1.
PT DOMAIN 102 140 EGF-LIKE 2.
PT DOMAIN 141 177 EGF-LIKE 3.
PT DOMAIN 179 215 EGF-LIKE 4.
PT DOMAIN 217 254 EGF-LIKE 5.
PT DOMAIN 256 292 EGF-LIKE 6.
PT DOMAIN 294 332 EGF-LIKE 7.
PT DOMAIN 334 370 EGF-LIKE 8.
PT DOMAIN 371 409 EGF-LIKE 9.
PT DOMAIN 411 449 EGF-LIKE 10.
PT DOMAIN 451 487 EGF-LIKE 11.
PT DOMAIN 489 525 EGF-LIKE 12.
PT DOMAIN 527 563 EGF-LIKE 13.
PT DOMAIN 565 600 EGF-LIKE 14.
PT DOMAIN 602 638 EGF-LIKE 15.
PT DOMAIN 640 675 EGF-LIKE 16.
PT DOMAIN 677 713 EGF-LIKE 17.
PT DOMAIN 715 750 EGF-LIKE 18.
PT DOMAIN 752 788 EGF-LIKE 19.
PT DOMAIN 790 826 EGF-LIKE 20.
PT DOMAIN 828 866 EGF-LIKE 21.
PT DOMAIN 868 904 EGF-LIKE 22.
PT DOMAIN 906 942 EGF-LIKE 23.
PT DOMAIN 944 980 EGF-LIKE 24.
PT DOMAIN 982 1018 EGF-LIKE 25.
PT DOMAIN 1020 1056 EGF-LIKE 26.
PT DOMAIN 1058 1094 EGF-LIKE 27.
PT DOMAIN 1096 1142 EGF-LIKE 28.
PT DOMAIN 1144 1180 EGF-LIKE 29.
PT DOMAIN 1182 1218 EGF-LIKE 30.
PT DOMAIN 1220 1264 EGF-LIKE 31.
PT DOMAIN 1266 1304 EGF-LIKE 32.
PT DOMAIN 1306 1346 EGF-LIKE 33.
PT DOMAIN 1347 1383 EGF-LIKE 34.
PT DOMAIN 1386 1424 EGF-LIKE 35.
PT REPEAT 1441 1478 LIN/NOTCH 1.
PT REPEAT 1479 1520 LIN/NOTCH 2.
PT REPEAT 1521 1560 LIN/NOTCH 3.
PT REPEAT 1876 1919 ANK 1.
PT REPEAT 1924 1953 ANK 2.
PT REPEAT 1957 1987 ANK 3.
PT REPEAT 1991 2020 ANK 4.
PT REPEAT 2024 2053 ANK 5.
PT REPEAT 2057 2086 ANK 6.
PT DISULFID 22 35 BY SIMILARITY.
PT DISULFID 29 45 BY SIMILARITY.

```

FT DISULFID 47 56 BY SIMILARITY.  
FT DISULFID 62 74 BY SIMILARITY.  
FT DISULFID 68 87 BY SIMILARITY.  
FT DISULFID 89 98 BY SIMILARITY.  
FT DISULFID 106 117 BY SIMILARITY.  
FT DISULFID 111 128 BY SIMILARITY.  
FT DISULFID 130 139 BY SIMILARITY.  
FT DISULFID 145 156 BY SIMILARITY.  
FT DISULFID 150 165 BY SIMILARITY.  
FT DISULFID 167 176 BY SIMILARITY.  
FT DISULFID 183 194 BY SIMILARITY.  
FT DISULFID 188 203 BY SIMILARITY.  
FT DISULFID 205 214 BY SIMILARITY.  
FT DISULFID 221 232 BY SIMILARITY.  
FT DISULFID 226 242 BY SIMILARITY.  
FT DISULFID 244 253 BY SIMILARITY.  
FT DISULFID 260 271 BY SIMILARITY.  
FT DISULFID 265 280 BY SIMILARITY.  
FT DISULFID 282 291 BY SIMILARITY.  
FT DISULFID 298 311 BY SIMILARITY.  
FT DISULFID 305 320 BY SIMILARITY.  
FT DISULFID 332 349 BY SIMILARITY.  
FT DISULFID 338 358 BY SIMILARITY.  
FT DISULFID 343 369 BY SIMILARITY.  
FT DISULFID 360 375 BY SIMILARITY.  
FT DISULFID 375 386 BY SIMILARITY.  
FT DISULFID 380 397 BY SIMILARITY.  
FT DISULFID 399 408 BY SIMILARITY.  
FT DISULFID 415 428 BY SIMILARITY.  
FT DISULFID 422 437 BY SIMILARITY.  
FT DISULFID 439 448 BY SIMILARITY.  
FT DISULFID 455 466 BY SIMILARITY.  
FT DISULFID 460 475 BY SIMILARITY.  
FT DISULFID 477 486 BY SIMILARITY.  
FT DISULFID 493 504 BY SIMILARITY.  
FT DISULFID 498 513 BY SIMILARITY.  
FT DISULFID 515 524 BY SIMILARITY.  
FT DISULFID 531 542 BY SIMILARITY.  
FT DISULFID 536 551 BY SIMILARITY.  
FT DISULFID 553 562 BY SIMILARITY.  
FT DISULFID 569 579 BY SIMILARITY.  
FT DISULFID 574 588 BY SIMILARITY.  
FT DISULFID 590 599 BY SIMILARITY.  
FT DISULFID 606 617 BY SIMILARITY.  
FT DISULFID 611 626 BY SIMILARITY.  
FT DISULFID 628 637 BY SIMILARITY.  
FT DISULFID 644 654 BY SIMILARITY.  
FT DISULFID 649 663 BY SIMILARITY.  
FT DISULFID 665 674 BY SIMILARITY.  
FT DISULFID 681 692 BY SIMILARITY.  
FT DISULFID 686 701 BY SIMILARITY.  
FT DISULFID 703 712 BY SIMILARITY.  
FT DISULFID 719 729 BY SIMILARITY.  
FT DISULFID 724 738 BY SIMILARITY.  
FT DISULFID 740 749 BY SIMILARITY.  
FT DISULFID 756 767 BY SIMILARITY.  
FT DISULFID 761 776 BY SIMILARITY.  
FT DISULFID 778 787 BY SIMILARITY.  
FT DISULFID 794 805 BY SIMILARITY.  
FT DISULFID 799 814 BY SIMILARITY.  
FT DISULFID 816 825 BY SIMILARITY.  
FT DISULFID 832 843 BY SIMILARITY.  
FT DISULFID 837 854 BY SIMILARITY.  
FT DISULFID 856 865 BY SIMILARITY.  
FT DISULFID 872 883 BY SIMILARITY.  
FT DISULFID 877 892 BY SIMILARITY.  
FT DISULFID 894 903 BY SIMILARITY.  
FT DISULFID 910 921 BY SIMILARITY.  
FT DISULFID 915 930 BY SIMILARITY.  
FT DISULFID 932 941 BY SIMILARITY.  
FT DISULFID 986 997 BY SIMILARITY.  
FT DISULFID 991 1006 BY SIMILARITY.  
FT DISULFID 1008 1017 BY SIMILARITY.

FT DISULFID 1024 1035 BY SIMILARITY.  
FT DISULFID 1029 1044 BY SIMILARITY.  
FT DISULFID 1046 1055 BY SIMILARITY.  
FT DISULFID 1062 1073 BY SIMILARITY.  
FT DISULFID 1067 1082 BY SIMILARITY.  
FT DISULFID 1084 1093 BY SIMILARITY.  
FT DISULFID 1100 1121 BY SIMILARITY.  
FT DISULFID 1115 1130 BY SIMILARITY.

Query Match 9.8%; Score 148.5; DB 1; Length 2524;  
Best Local Similarity 22.5%; Pred. No. 0.0044;  
Matches 73; Conservative 32; Mismatches 102; Indels 117; Gaps 19;

Qy 34 MHPNVSQCGGGATCSYNGCLSKCPRLPALERIGH-----KQIGV 76  
Db 869 MNECVNPPCRNG-ATCQNTNGSYKCNCKPGTGRNCEMDIDDCQPNPCHNGSGSCDGINW 927  
Qy 77 CLSSCPGYYGTR-YPDINKCTK--CK--ADCDTCFKNFCTCKSGFY-LHL----- 123  
Db 928 PFCNCPAGFRGPRCEEDINECASNPCKNGANCTCVNSYTC-CQPGSGIHCSNTPD 986  
Qy 124 -----GKCLD-----NCPGLEAN--NHTM-ECVSI----- 146  
Db 987 TESSCFNGGTCIDGINTFTCCPPGFTGSGYQHDINECDKPKLNGGTCDSDSYGTYKTC 1046  
Qy 147 -----VHCE-VSEWNPWSPCTKKG-----TCGFKG-----TETVRE 179  
Db 1047 PQGYTGLANQNLVRWCSDSPCKGCKWQNNFYRCBCKSGMTGVYCDVPVSVCEVAARQ 1106  
Qy 180 ---LIQHSAGKMLCPPTNETRKTQVQKGERGKGRKRK-KPNK----- 226  
Db 1107 QGVDIVHLCRNSGMCVDTGNTHFC-----RCQAGYTSYCEQVDECSNFCQNGATCTD 1161  
Qy 227 ---GESKEAIPDSKSLSSKEIPE 247  
Db 1162 YLGYSCECVAGYGVNCSBEINE 1185

RESULT 14  
NTC2 RAT STANDARD; PRT; 2471 AA.  
AC Q9QW30;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).  
GN NOTCH2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Brain;  
RC MEDLINE=93202015; PubMed=1295745;  
RX Weinmaster G., Roberts V.J., Lemke G.;  
RA "Notch2: a second mammalian Notch gene.";  
RL Development 116:931-941(1992).  
RN [2]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=21311789; PubMed=11438922;  
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;  
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple  
functional roles for the Notch-DSL signaling system during brain  
development.";  
RT J. Comp. Neurol. 436:167-181(2001).  
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands  
Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs. May play an essential role in postimplantation

development, probably in some aspect of cell specification and/or differentiation (By similarity).

-!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(SC) which are probably linked by disulfide bonds (By similarity).

-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.

-!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.

-!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.

-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(SC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: Belongs to the NOTCH family.

-!- SIMILARITY: Contains 35 EGF-like domains.

-!- SIMILARITY: Contains 2 Lin/Notch repeats.

-!- SIMILARITY: Contains 6 ANK repeats.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M93661; AAK13558.1; --

PIR; A49128; A49128.

HSSP; P00743; 1CCF.

InterPro: IPR002110; ANK.

InterPro: IPR000152; Asx\_hydroxyl\_S.

InterPro: IPR000742; EGF\_2.

InterPro: IPR001881; EGF\_Ca.

InterPro: IPR001438; EGF\_II.

InterPro: IPR006209; EGF\_like.

InterPro: IPR002049; Laminin\_EGF.

InterPro: IPR008297; Notch.

InterPro: IPR008000; Notch\_dom.

Pfam; PF00023; ank; 6.

Pfam; PF00008; EGF; 35.

Pfam; PF00066; notch; 2.

Pfam; PF002279; Notch; 1.

PRINTS; PR00010; EGFBL00D.

PRINTS; PR00011; EGF\_LAMININ.

PRINTS; PR01452; NOTCH.

SMART; SM00248; ANK; 6.

SMART; SM00179; EGF\_CA; 24.

SMART; SM00004; NL; 2.

PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.

PROSITE; PS50088; ANK\_REPEAT; 4.

PROSITE; PS00010; ASX\_HYDROXYL; 22.

PROSITE; PS00022; EGF\_1; 34.

PROSITE; PS01186; EGF\_2; 26.

PROSITE; PS50026; EGF\_3; 35.

PROSITE; PS01187; EGF\_CA; 22.

Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

SIGNAL 1 25

CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.

CHAIN	FT	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
CHAIN	FT	1697	2471	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY)
DOMAIN	FT	26	1677	EXTRACELLULAR (POTENTIAL).
TRANSMEM	FT	1678	1698	POTENTIAL.
DOMAIN	FT	1699	2471	CYTOPLASMIC (POTENTIAL).
DOMAIN	FT	26	63	EGF-LIKE 1.
DOMAIN	FT	64	102	EGF-LIKE 2.
DOMAIN	FT	105	143	EGF-LIKE 3.
DOMAIN	FT	144	180	EGF-LIKE 4.
DOMAIN	FT	182	219	EGF-LIKE 5.
DOMAIN	FT	221	258	EGF-LIKE 6.
DOMAIN	FT	260	296	EGF-LIKE 7.
DOMAIN	FT	298	336	EGF-LIKE 8.
DOMAIN	FT	338	374	EGF-LIKE 9.
DOMAIN	FT	375	413	EGF-LIKE 10.
DOMAIN	FT	415	454	EGF-LIKE 11.
DOMAIN	FT	456	492	EGF-LIKE 12.
DOMAIN	FT	494	530	EGF-LIKE 13.
DOMAIN	FT	532	568	EGF-LIKE 14.
DOMAIN	FT	570	605	EGF-LIKE 15.
DOMAIN	FT	607	643	EGF-LIKE 16.
DOMAIN	FT	645	680	EGF-LIKE 17.
DOMAIN	FT	682	718	EGF-LIKE 18.
DOMAIN	FT	720	755	EGF-LIKE 19.
DOMAIN	FT	757	793	EGF-LIKE 20.
DOMAIN	FT	795	831	EGF-LIKE 21.
DOMAIN	FT	833	871	EGF-LIKE 22.
DOMAIN	FT	873	909	EGF-LIKE 23.
DOMAIN	FT	911	947	EGF-LIKE 24.
DOMAIN	FT	949	985	EGF-LIKE 25.
DOMAIN	FT	987	1023	EGF-LIKE 26.
DOMAIN	FT	1025	1061	EGF-LIKE 27.
DOMAIN	FT	1063	1099	EGF-LIKE 28.
DOMAIN	FT	1101	1147	EGF-LIKE 29.
DOMAIN	FT	1149	1185	EGF-LIKE 30.
DOMAIN	FT	1187	1223	EGF-LIKE 31.
DOMAIN	FT	1225	1262	EGF-LIKE 32.
DOMAIN	FT	1264	1302	EGF-LIKE 33.
DOMAIN	FT	1304	1343	EGF-LIKE 34.
DOMAIN	FT	1374	1412	EGF-LIKE 35.
DOMAIN	FT	1645	1648	POLY-ALA.
DOMAIN	FT	1994	1997	POLY-SER.
DOMAIN	FT	2426	2429	POLY-GLY.
REPEAT	FT	2446	2451	LIN/NOTCH 1.
REPEAT	FT	1420	1456	LIN/NOTCH 2.
REPEAT	FT	1503	1535	ANK 1.
REPEAT	FT	1827	1871	ANK 2.
REPEAT	FT	1876	1905	ANK 3.
REPEAT	FT	1909	1939	ANK 4.
REPEAT	FT	1943	1972	ANK 5.
REPEAT	FT	1976	2005	ANK 6.
REPEAT	FT	2009	2036	BY SIMILARITY.
DISULFID	FT	28	41	BY SIMILARITY.
DISULFID	FT	35	51	BY SIMILARITY.
DISULFID	FT	53	62	BY SIMILARITY.
DISULFID	FT	68	79	BY SIMILARITY.
DISULFID	FT	73	90	BY SIMILARITY.
DISULFID	FT	92	101	BY SIMILARITY.
DISULFID	FT	109	121	BY SIMILARITY.
DISULFID	FT	115	131	BY SIMILARITY.
DISULFID	FT	133	142	BY SIMILARITY.
DISULFID	FT	148	159	BY SIMILARITY.
DISULFID	FT	153	168	BY SIMILARITY.
DISULFID	FT	170	179	BY SIMILARITY.
DISULFID	FT	186	198	BY SIMILARITY.
DISULFID	FT	192	207	BY SIMILARITY.
DISULFID	FT	209	218	BY SIMILARITY.
DISULFID	FT	225	236	BY SIMILARITY.
DISULFID	FT	230	246	BY SIMILARITY.
DISULFID	FT	248	257	BY SIMILARITY.
DISULFID	FT	264	275	BY SIMILARITY.

```

FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.

Query Match 9.7%; Score 147.5; DB 1; Length 2471;
Best Local Similarity 24.8%; Pred. No. 0.005;
Matches 54; Conservative 23; Mismatches 52; Indels 89; Gaps 13;

QY 31 QRMHNVSGCGGATCSDYNGCTSCPKRLPFALERIGKQIGVCLLSCPSGYGTRY 90
DB 947 QTDNECLSPCKNG-GTCSDYVNSYTC-----TCPAGFHGVHC 984
QY 91 P-DINKCTCKADCDTCENKFC-----TKCKSGFLYHLGK 125
DB 985 ENNIDETE-----SSCFNGGTGVDGINSFSLCPVGFTGPFCLHDINECSSNPCLNSGT 1039
QY 126 CLD-----NCEGLEANNHTMECVSIHV-CEVSENNPWSPTCKGKTKGPKRGITR 177
DB 1040 CVDGLGYRCTCLGYTGKN-----CQTLVNLG-----SP-SPCKNG-TCA----- 1079
QY 178 REIIQHPKAGNLCPTNETRKTCTVQRKCC-----QNG 210
DB 1080 ----QEKARPRCLCPGWDGAYCDVLNVSKAAALQNG 1113

RESULT 15
BL14_CABEL
ID BL14_CABEL STANDARD; PRT; 943 AA.
AC P51559; O44762; O44763; O44764; O44765; O44766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoprotease bli-4 precursor (EC 3.4.21.-) (Blisterase) (Blistered
cuticle protein 4).
GN BLI-4 OR KPC-4 OR K04F10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN BLI-4_CABEL
SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=95293228; PubMed=7774813;
RA Thacker C., Peters K.W., Srayko M., Rose A.M.;
RT distinct kex2/subtilisin-like endoproteases essential for early
development and adult morphology.;
RL Genes Dev. 9:956-971(1995).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RA Latreille P., Wameley P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP MUTAGENESIS.
```

```

RX MEDLINE=20363723; PubMed=10903434;
RA Thacker C., Srayko M., Rose A.M.;
RT "Mutational analysis of bli-4/kpc-4 reveals critical residues required
for proprotein convertase function in C. elegans.";
RL Gene 252:15-25(2000).
CC -1- FUNCTION: The Kex2/subtilisin-like proteinase activity of this
enzyme is required for the normal production of adult cuticle.
There is functional redundancy between the isoforms. Vital for
embryonic and larval development.
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=5;
Name=D; Synonyms=d;
IsoId=P51559-1; Sequence=Displayed;
Name=A; Synonyms=a;
IsoId=P51559-2; Sequence=VSP_005416, VSP_005419;
Name=B; Synonyms=b;
IsoId=P51559-3; Sequence=VSP_005418, VSP_005421;
Name=C; Synonyms=c;
IsoId=P51559-4; Sequence=VSP_005422, VSP_005423;
Name=E;
IsoId=P51559-5; Sequence=VSP_005417, VSP_005420;
Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: In larvae and adults, expressed in all
hypodermal cells, vulva and ventral nerve cords.
CC -1- DEVELOPMENTAL STAGE: Expression starts at embryo two-fold stage
through to adults.
CC -1- SIMILARITY: Belongs to peptidase family S8. Purin subfamily.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 651.
CC -1- CAUTION: Ref.2 (AAB96754 and AAB96757) sequence differs from that
shown due to erroneous gene model prediction.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.iesb-sib.ch/announce/
or send an email to license@iesb-sib.ch).
CC -----
EMBL; L29438; AAA98750.1; ALT_FRAME.
EMBL; L29439; AAA98751.1; ALT_FRAME.
EMBL; L29440; AAA98752.1; ALT_FRAME.
EMBL; AF039719; AAB96753.1; -.
EMBL; AF039719; AAB96754.1; ALT_SEQ.
EMBL; AF039719; AAB96755.1; -.
EMBL; AF039719; AAB96756.1; -.
EMBL; AF039719; AAB96757.1; ALT_SEQ.
MEROPS; S08.031; -.
WormPep; K04F10.4a; CE11728.
WormPep; K04F10.4b; CE11730.
WormPep; K04F10.4c; CE11732.
WormPep; K04F10.4d; CE11734.
WormPep; K04F10.4e; CE11736.
GO; GO:0016021; C: integral to membrane; NAS.
GO; GO:0004352; P: serine-type endopeptidase activity; NAS.
GO; GO:0007592; P: cuticle biosynthesis (sensu Invertebrata); IMP.
InterPro; IPR006212; Furin repeat.
InterPro; IPR009030; Grow fac recep.
InterPro; IPR000209; Peptidase S8.
InterPro; IPR002884; Peptidase_S8B.
InterPro; IPR009020; Protease_inhib.
Pfam; PF01483; P: protease; 1.
Pfam; PF00082; Peptidase_S8; 1.
PRINTS; PR00723; SUBTILISIN.
ProDom; PD000717; P domain; 1.
SMART; SM00261; FU; 3.
PROSITE; PS00136; SUBTILASE_ASP; 1.
PROSITE; PS00137; SUBTILASE_HIS; 1.
PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Calcium-binding; Zymogen;
KW Transmembrane; Signal; Alternative splicing.
FT SIGNAL 1 20
POTENTIAL.
```

Search completed: June 29, 2004, 17:00:16  
Job time : 12.2828 secs

Query Match 9.6%; Score 145; DB 1; Length 943;  
Best Local Similarity 25.6%; Pred. No. 0.0031;  
Matches 51; Conservative 21; Mismatches 77; Indels 50; Gaps 11;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:55:24 ; Search time 34.7876 Seconds  
(without alignments)  
2476.067 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISLWFLINPMEYI.....QQKKRVQDKSVSVTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1505	99.3	272	4 Q9BX14	Q9bx14 homo sapien
2	1452	95.8	292	4 Q96K87	Q96k87 homo sapien
3	1153	76.1	217	11 Q9CSB2	Q9csb2 mus musculu
4	1064	70.2	224	11 Q8BVW2	Q8bvw2 mus musculu
5	644	42.5	255	11 Q92132	Q92132 mus musculu
6	626.5	41.3	236	4 Q8N7L5	Q8n7l5 homo sapien
7	576.5	38.0	243	11 Q8BFU0	Q8bfu0 mus musculu
8	472	31.1	224	4 Q9UGB2	Q9ugb2 homo sapien
9	459.5	30.3	176	4 Q8N6X6	Q8n6x6 homo sapien
10	417.5	27.5	152	11 Q7TPX3	Q7tpx3 mus musculu
11	283.5	18.7	138	11 Q8BJ73	Q8bj73 mus musculu
12	192	12.7	129	5 Q26489	Q26489 apodoptera
13	179.5	11.8	913	13 Q8AY18	Q8ay18 rana esculu
14	177	11.7	1101	5 Q964D2	Q964d2 entamoeba h
15	176.5	11.6	1376	5 Q8SZS2	Q8szs2 drosophila
16	176	11.6	1074	5 Q964D1	Q964d1 entamoeba h

17	172	11.3	915	11 Q91VK0	Q91vk0 mus musculu
18	167.5	11.0	296	11 Q35171	Q35171 mus musculu
19	167.5	11.0	826	11 Q8CFZ2	Q8cfz2 mus musculu
20	167.5	11.0	932	11 Q62030	Q62030 mus musculu
21	163	10.8	3869	5 Q86PQ3	Q86pq3 cryptospori
22	162	10.7	913	4 Q96EP4	Q96ep4 homo sapien
23	161	10.6	440	5 Q18003	Q18003 caenorhabdi
24	159	10.5	503	5 Q9U018	Q9u018 giardia lam
25	157.5	10.4	803	13 Q42114	Q42114 brachydanio
26	156	10.3	898	5 Q76822	Q76822 branchiosco
27	155	10.2	808	13 Q42113	Q42113 brachydanio
28	153.5	10.1	1362	13 Q9PVZ4	Q9pvz4 xenopus lae
29	151	10.0	238	5 Q76510	Q76510 cryptospori
30	149.5	9.9	4007	4 Q86XX4	Q86xx4 homo sapien
31	149	9.8	1371	11 Q9QVW4	Q9qvW4 rattus sp.
32	148	9.8	548	5 Q9GQ45	Q9gq45 giardia lam
33	148	9.8	4010	11 Q80T14	Q80t14 mus musculu
34	147.5	9.7	802	13 Q9W770	Q9w770 gallus gall
35	146.5	9.7	365	11 Q924Y6	Q924y6 rattus norv
36	146.5	9.7	807	4 Q9HCB6	Q9hcb6 homo sapien
37	145	9.6	724	4 Q94862	Q94862 homo sapien
38	145	9.6	807	4 Q8NCD7	Q8ncd7 homo sapien
39	144.5	9.5	660	5 Q23832	Q23832 cryptospori
40	144.5	9.5	2189	5 Q9BI05	Q9bi05 eimeria ten
41	144	9.5	807	6 Q9GLX9	Q9glx9 bos taurus
42	143.5	9.5	220	11 Q99KR2	Q99kr2 mus musculu
43	143.5	9.5	402	11 Q8K2Q8	Q8k2q8 mus musculu
44	143.5	9.5	807	11 Q8VCC9	Q8vcc9 mus musculu
45	143	9.4	435	5 Q9GQ41	Q9gq41 giardia lam

ALIGNMENTS

RESULT 1

Q9BX14 ID Q9BX14 PRELIMINARY; PRT; 272 AA.  
AC Q9BX14;  
DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Thrombospondin.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mao Y., Xie Y., Zhou Z., Zhao W., Wang S., Huang Y., Wang S.,  
Tang R., Chen X., Wu C.,  
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUS=Placenta;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR ENBL; AP251057; AAK34947.1; -;  
DR ENBL; BC022367; AAK22367.1; -;  
DR Genew; HGNC:20866; THSD2.  
DR InterPro; IPR006212; Furin\_repeat.  
DR InterPro; IPR009030; Grow\_fac\_recep.  
DR InterPro; IPR000884; TSPI.  
DR Pfam; PF00090; tspl\_1; 1.  
DR SMART; SM00261; FU; 2.  
DR SMART; SM00209; TSPI; 1.  
DR PROSITE; PSS0092; TSPI; 1.  
SQ SEQUENCE 272 AA; 30928 MW; CACAEC6B7E781189 CRC64;

Query Match 99.3%; Score 1505; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. No. 4.3e-130; Indels 0;

Matches 271; Conservative 0; Mismatches 0; Gaps 0;

QY 3 HLRISLWFLINPMEYIQQKKRVQDKSVSVTVH 62

```
Db 2 HLLRISWLFILNFMFYIGSNASRGRQRMRHNPVSGQGGCATCSDYNGCLSCPKRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCDTCFNNKFTCKKSGPYLH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCDTCFNNKFTCKKSGPYLH 121
Qy 123 LGKCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTRVREIIQ 182
Db 122 LGKCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTRVREIIQ 181
Qy 183 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKKPKNGESKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKKPKNGESKEAIPDSKSLESS 241
Qy 243 KEIPEORENKQOKKRVQDKQKSVSVSTVH 273
Db 242 KEIPEORENKQOKKRVQDKQKSVSVSTVH 272

RESULT 2
Q96K87 PRELIMINARY; PRT; 292 AA.
AC Q96K87;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14440.
OS Homo sapiens (Human)
OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiwa S., Komai P., Hara R., Takeuchi K.,
RA Arita M., Nabekura Y., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakanatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RA "WEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027346; BAB55051.1; -.
DR InterPro; IPR006212; Purin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
KM Hypothetical protein.
SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3DA6F8 CRC64;

Query Match 95.8%; Score 1452; DB 4; Length 292;
Best Local Similarity 99.2%; Pred. No. 3.3e-125;
Matches 262; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 HLLRISWLFILNFMFYIGSNASRGRQRMRHNPVSGQGGCATCSDYNGCLSCPKRL 62
Db 2 HLLRISWLFILNFMFYIGSNASRGRQRMRHNPVSGRQGGCATCSDYNGCLSCPKRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCDTCFNNKFTCKKSGPYLH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCDTCFNNKFTCKKSGPYLH 121
Qy 123 LGKCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTRVREIIQ 182
Db 122 LGKCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTRVREIIQ 181
Qy 183 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKKPKNGESKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKKPKNGESKEAIPDSKSLESS 241
```

```
Qy 243 KEIPEORENKQOKKRVQDKQKSVSVSTVH 266
Db 242 KEIPEORENKQOKKRVQDKQKSVSVSTVH 265

RESULT 3
Q9CSB2 PRELIMINARY; PRT; 217 AA.
AC Q9CSB2;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE 2810459H04RIK protein (Fragment).
GN THSD2 OR 2810459H04RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guncionich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszynski-Borja A., Yoshida K., Haegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK013366; BAB28811.1; -.
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR006212; Purin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 217
SQ SEQUENCE 217 AA; 24304 MW; 0DCF938B9FB3FB7 CRC64;

Query Match 76.1%; Score 1153; DB 11; Length 217;
Best Local Similarity 93.1%; Pred. No. 6.7e-98;
Matches 201; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 3 HLLRISWLFILNFMFYIGSNASRGRQRMRHNPVSGQGGCATCSDYNGCLSCPKRL 62
Db 2 HLLRISWLFILNFMFYIGSNASRGRQRMRHNPVSGQGGCATCSDYNGCLSCPKRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCDTCFNNKFTCKKSGPYLH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCDTCFNNKFTCKKSGPYLH 121
Qy 123 LGKCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTRVREIIQ 182
Db 122 LGKCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGPKRGTRVREIIQ 181
Qy 183 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKKPKNGESKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKKPKNGESKEAIPDSKSLESS 241
```



```

RESULT 4
Q8BVW2 PRELIMINARY; PRT; 224 AA.
AC Q8BVW2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thrombospondin homolog.
GN THSD2 OR 2810459H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK076308; BAC36296.1; -.
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR SMART; SM00261; FU; 2.
DR SQU SEQUENCE 224 AA; 25398 MW; BC13B083497CFEE3 CRC64;

Query Match 70.2%; Score 1064; DB 11; Length 224;
Best Local Similarity 84.6%; Pred. No. 1e-89;
Matches 187; Conservative 9; Mismatches 25; Indels 0; Gaps 0;

QY 3 HRLISWLFILNFMFYIGSQNASRGRQRMRHPNVSGQGGCATCDYNGCLSKPRL 62
DB 2 HRLISWLFILNFMFYIGSQNASRGRQRMRHPNVSGQGGCATCDYNGCLSKPRL 61
QY 63 FVALERIGMKQIGVCLSSCPGSGYGYTPYDINKTKCKADCTCFNKGCTCKSGYLH 122
DB 62 FVLERIGMKQIGVCLSSCPGSGYGYTPYDINKTKCKADCTCFNKGCTCKSGYLH 121
QY 123 LGKCLDNCPEGLANNHTECVSIHCEVSEWNPSPCTKKGKTCGPKGTETRVREIIQ 182
DB 122 LGKCLDNCPEGLANNHTECVSIHCEVSEWNPSPCTKKGKTCGPKGTETRVREIIQ 181
QY 183 HPSAGNLCPTNTRKCTVQRKKCKQKGRGKGRERKRX 223
DB 182 HPSAGNLCPTNTRKCTVQRKKCKQKGRGKGRERKRX 222

RESULT 5
Q92132 PRELIMINARY; PRT; 265 AA.
AC Q92132;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thrombospondin type 1 domain.
GN RSPONDIN OR R-SPONDIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;
RT "r-spondin, a novel thrombospondin type 1 domain gene, expressed in
RT the dorsal neural tube."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016768; BAA75640.1; -.
DR MGD; MGI:2183426; Rspndin.

```

```

DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR Pfam; PF00084; TSP1.
DR SMART; SM00261; FU; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
SQ SEQUENCE 265 AA; 29331 MW; FFE8964743F5963 CRC64;

Query Match 42.5%; Score 644; DB 11; Length 265;
Best Local Similarity 46.5%; Pred. No. 4.1e-51;
Matches 119; Conservative 36; Mismatches 77; Indels 24; Gaps 6;

QY 6 LISWLFILNFMFYIGSQNASRGRQRMRHPNVSGQGGCATCDYNGCLSKPRLPFA 65
DB 11 VLSTWTHIA-----VGSRG-I-KGRQRRI-SABGSOACAKGCELVNGLCKSPKLFIL 63
QY 66 LERIGMKQIGVCLSSCPGSGYGYTPYDINKTKCKAD-CDTCFNKGCTCKSGYLHIG 124
DB 64 LERNDIRQVGVCLPSCPFGYFDARNPDWNTKICKIEHCEACFPHNFCCKQBALYLHG 123
QY 125 KCLDNCPEGLANNHTECVSIHCEVSEWNPSPCTKKGKTCGPKGTETRVREIIQHP 184
DB 124 RYPACPEGSTAANSTMECGSPACQCESEWNPSPCTKKGKTCGPKGTETRVREIIQHP 183
QY 185 SAKGNLCPTNTRKCTVQRKKCKQKGRGKKG-----RERKPKPKNGKESKEAIPDSKSL 240
DB 184 GGDHTTCSDTKTKCTVRETTPCPBGQKRRKGGRHNNHAPRKNKSK--PRSNS-- 239
QY 241 SKSEIPQRNKQOK 256
DB 240 -----RRHKGQQ 247

RESULT 6
Q8N7L5 PRELIMINARY; PRT; 236 AA.
AC Q8N7L5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ40906.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Oshina A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito H., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wogatusuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098225; BAC05263.1; -.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; TSP1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
KW Hypothetical protein.
SQ SEQUENCE 236 AA; 25942 MW; 8D03803127EC5678 CRC64;

Query Match 41.3%; Score 626.5; DB 4; Length 236;
Best Local Similarity 48.1%; Pred. No. 1.5e-49;
Matches 111; Conservative 36; Mismatches 69; Indels 15; Gaps 3;

```

```
QY 33 RMHPNVSCQCGCATCSYNGCLSCPKPRLFPALERIGMKQIGVCLSSCPGYYGTRYD 92
DB 4 RVSAEGSQACAKGCELCSEVNGCLCKSPKLFILLERNDIRQVGLCLPSCPYPGDFARNPD 63
QY 93 INKCTKAD-CDTCFNKPNCTCKSGFYHLGLKCLDNCPEGLAANNHTMECVSIHVCEV 151
DB 64 MNKCIACKLEHCEACFSHFNCTCKSGFYHLGLKCLDNCPEGLAANNHTMECVSIHVCEV 123
QY 152 SEWNPMSPTCKGKTCGFKGTTRVRIIHOHPSAKGNLCPPNTRKCTVQKCKQKE 211
DB 124 SEWNPMSPTCKGKTCGFKGTTRVRIIHOHPSAKGNLCPPNTRKCTVQKCKQKE 183
QY 212 RGKKGHRERKPNKG-----ESKEAIPDSKSLSSKEIPEQRENKQOQKR 258
DB 184 KRRKGQGRRENARNRLARKEBAGAGSR-----RRKQOQOQOQO 224

RESULT 7
Q8BFU0
ID Q8BFU0 PRELIMINARY; PRT; 243 AA.
AC Q8BFU0
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical thrombospondin type I repeat.
GN 261002F08RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bye, and Hippocampus;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK049891; BAC33974.1; -.
DR EMBL; AK087485; BAC39893.1; -.
DR MGD; MGI:1922667; 261002F08RIK.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSPI.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
KW Hypothetical protein.
SQ SEQUENCE 243 AA; 28275 MW; ED76A08D61012ED7 CRC64;

Query Match 38.0%; Score 576.5; DB 11; Length 243;
Best Local Similarity 45.0%; Pred. No. 5.9e-45;
Matches 108; Conservative 39; Mismatches 78; Indels 15; Gaps 5;

QY 6 LISWLIILNPMFYIGSONASRGRRQPRMHPNVSCQCGCATCSYNGCLSCPKPRLFP 65
DB 5 LPSFALLIILNPMFYIGSONASRGRRQPRMHPNVSCQCGCATCSYNGCLSCPKPRLFP 62
QY 66 LERIGHKQIGVCLSSCPGYYGTRYDIPINKCTKAD-CDTCFNKPNCTCKSGFYHLGL 124
DB 63 LRREGHRYQVECLHSCPSGYGHRAPDMNRCARCIENCDSCFSKDFCTCKVGYLHRG 122
QY 125 KCLDNCPEGLAANNHTMECVSIHVCEWNPMSPTCKGKTCGFKGTTRVRIIHOH 184
DB 123 RCFDECPDGPAPLDETWECVE--GCEVHSEWGTCSRNRRTCGFKWGLETRQIVKPP 180
QY 185 SAKGNLCPPNTRKCTVQKCKQKGRGKGRERKKNKESKAI PDSKSLSSKE 244
DB 181 AKDTIPCTTAESRRRCQAMRHCPGGKRTTPAKEKRNKKRR-----KLITERAQE 230

RESULT 8
```

```
Q9UGB2
ID Q9UGB2 PRELIMINARY; PRT; 224 AA.
AC Q9UGB2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DJ824F16.3 (Novel protein similar to mouse thrombospondin type 1
DE domain protein R-spondin) (Fragment).
GN DJ824F16.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Blakey S.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL050325; CAB65783.3; -.
DR Genew; HGNC:16175; C20orf182.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSPI.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
FT NON TER 224
SQ SEQUENCE 224 AA; 25042 MW; 97D26AD34CDBF812 CRC64;

Query Match 31.1%; Score 472; DB 4; Length 224;
Best Local Similarity 43.3%; Pred. No. 2.1e-35;
Matches 87; Conservative 36; Mismatches 70; Indels 8; Gaps 5;

QY 29 RRQRMRHPNVSCQCGCATCSYNGCLSCPKPRLFPALERIGMKQIGVCLSSCPGYYGT 88
DB 22 RRRKQVGTGLGGNCT-GCIICSENGCSTCQQLFLFIRREGINQYKGLHDCPPGFGI 80
QY 89 RYDPINKCTKADCDTCFNKPNCTCKSGFYHLGLKCLDNCPEGLAANNHTMECVSIHV 148
DB 81 RGQEVNRCKCGATCSYNGCLSCPKPRLFPALERIGMKQIGVCLSSCPGYYGT 138
QY 149 CVSEWNPMSPTCKGKTCGFKGTTRVRIIHOHPSAKGNLCPPNTRKCTVQKCKQ 208
DB 139 CHLPGWGSPTCKGKTCGFKGTTRVRIIHOHPSAKGNLCPPNTRKCTVQKCKQ 196
QY 209 KGERG---KGRERKKNK 226
DB 197 PGRSPQKQKGRKRRPKDR 217

RESULT 9
Q8N6X6
ID Q8N6X6 PRELIMINARY; PRT; 176 AA.
AC Q8N6X6
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to putative.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lung;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027938; AAH27938.1; -.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSPI.
DR SMART; SM00261; FU; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
```

```
SQ SEQUENCE 176 AA; 20409 MW; 0F83CCE1B2F8CA85 CRC64;
Query Match 30.3%; Score 459.5; DB 4; Length 176;
Best Local Similarity 45.1%; Pred. No. 2.3e-34;
Matches 79; Conservative 30; Mismatches 53; Indels 13; Gaps 3;

QY 71 MKQIGVCLSSCPGSGYGYTRYPDINKTKCKAD-CDTCFKNKFNCTKCKSGFYHLGKCLDN 129
DB 1 MRQYGECLHSCPSGYGHRAPDMNRCARCIENCDSCFSDKDFCTCKCKVGYLHRCFDE 60

QY 130 CPBGLANNHTECVSIVHCEVSEWNPSPCTKKGKTCGPRGTRVRIIHPHSAKN 189
DB 61 CPDGPAPLBETMECV--GCEVGHWSGWTGSRNRTCGFKGLETRQIVKPKVDYTI 118

QY 190 LCPPTNETKCTVQKCKQGGKGRKRKPKNGESKEAIPDKSLSSEKE 244
DB 119 PCPTIABRRRCNMTMRHCPGGRTPKAKEKRNKKKKR-----KLIFRAQE 163

RESULT 10
Q7TPX3
ID Q7TPX3 PRELIMINARY; PRT; 152 AA.
AC Q7TPX3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 2610028F08Rik protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozney K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smalish U., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052844; AAHS2844.1; -.
SQ SEQUENCE 152 AA; 17658 MW; B74713789B285384 CRC64;
Query Match 27.5%; Score 417.5; DB 11; Length 152;
Best Local Similarity 52.3%; Pred. No. 1.4e-30;
Matches 79; Conservative 20; Mismatches 47; Indels 5; Gaps 4;

QY 6 LISWPIILNFWYIGSQVARSRCRRRRMHPNVSCQCGCATCSYNGCLSKCPPLPFA 65
DB 5 LPSFALIILNCMDYSQCG--GNRRRRKRA--SYVSNPICRGCLSKSDNGSCRCQCKLPPF 62

QY 66 LRIQGMQKIGVCLSSCPGSGYGYTRYPDINKTKCKAD-CDTCFKNKFNCTKCKSGFYHLG 124
```

```
DB 63 LRREKMQYGECLHSCPSGYGHRAPDMNRCARCIENCDSCFSDKDFCTCKCKVGYLHRC 122
QY 125 KCLDNCPBGLANNHTECVSIVHCEVSEWN 155
DB 123 RCFDRCPDGSPAPLDTEMECV--GCEVGHWS 151

RESULT 11
Q8BJ73
ID Q8BJ73 PRELIMINARY; PRT; 138 AA.
AC Q8BJ73;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DJ824F16.3
GN A930029KL9RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK020904; BAC25643.1; -.
DR MGD; MGI:1924467; A930029KL9RIK.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_Fac_recep.
DR SMART; SM00261; FU; 2.
SQ SEQUENCE 138 AA; 15172 MW; FEFD7D949279D5DD CRC64;
Query Match 18.7%; Score 283.5; DB 11; Length 138;
Best Local Similarity 42.9%; Pred. No. 2.5e-18;
Matches 45; Conservative 24; Mismatches 35; Indels 1; Gaps 1;

QY 29 RRORRHNPVSCQCGCATCSYNGCLSKCPPLPALERIGMKQIGVCLSSCPGSGYGT 88
DB 22 RRRQKAGTGLGNGCT--GCVCSENGCSTCQRLFLFIRREGIRQYKGVHDCPLGPGFI 80

QY 89 RYPDINKTKCKADCDTCFKNKFNCTKCKSGFYHLGKCLDNCPG 133
DB 81 RQGEANRCKKOGATCSYNGCLSKCPPLPALERIGMKQIGVCLSSCPGSGYGT 125

RESULT 12
Q26489
ID Q26489 PRELIMINARY; PRT; 1299 AA.
AC Q26489;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endoprotease furin.
GN Furin.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyridae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Cieplik M., Klenk H.;
RT "Cloning and functional characterization of FURIN from Spodoptera
RT frugiperda (sf9) cells."
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9.
CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
DR EMBL; Z68888; CAA93116.1; -.
```

DR InterPro: IPR006212; Purin\_repeat.  
DR InterPro: IPR009030; GroW\_fac\_recep.  
DR InterPro: IPR006210; IEGF.  
DR InterPro: IPR000209; Peptidase\_S8.  
DR InterPro: IPR002884; Peptidase\_S8B.  
DR InterPro: IPR009020; Protease\_Inhib.  
DR Pfam: PF00757; Purin-like; 1.  
DR Pfam: PF00082; Peptidase\_S8; 1.  
DR Pfam: PF01483; P\_protein; 1.  
DR PRINTS: PR00723; SUBTILISIN.  
DR PRODOM: PD000717; P\_domain; 1.  
DR SMART: SM00181; EGF; 4.  
DR SMART: SM00261; FU; 5.  
DR PROSITE: PS00190; CYTOCHROME C; 2.  
DR PROSITE: PS00136; SUBTILASE ASP; 1.  
DR PROSITE: PS00137; SUBTILASE HIS; 1.  
DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
KW Protease.  
SQ SEQUENCE 913 AA; 101864 MW; 6C9020632C47D9BE CRC64;  
  
Query Match 11.8%; Score 179.5; DB 13; Length 913;  
Best Local Similarity 29.4%; Pred. No. 5.9e-08;  
Matches 52; Conservative 22; Mismatches 66; Indels 37; Gaps 11;  
  
QY 44 GGC-ATCSDYNGC-----LSCKPRLPALERIGHKQICVCLSSCPSCGYGTRYPDINKCT 97  
DB 632 GPCDAECSDV-GCDGPGPHHCNCLHFFYK--AKNTRICVSDCPLGYPA---DKKCK 685  
  
QY 98 KCKADGDTCP--KNFCTCKSGFYLLH--LQKLDNCPGELANNHMTBVCVSIHVCEVSE 153  
DB 686 KCPPTCYCLGSRSDQSCSKSGYLNEETNSCTVNCDFGYLNDKNLC----- 735  
  
QY 154 WNPWSPCTKKGKTCGPKRGTRVREIIQHP-SAKGNLCPTTNET-RKCTVQRKKCQ 208  
DB 736 ---KXCHENCKIC----TSABICTECEGLSLQGSTCAVTCSDGRYSYSAVKKECE 783  
  
RESULT 14  
Q964D2 PRELIMINARY; PRT; 1101 AA.  
AC ID Q964D2 PRELIMINARY; PRT; 1101 AA.  
CD Q964D2  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Gal/GalNac lectin Ig1.  
GN IGL1.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
[1]  
RN R1  
RP SEQUENCE FROM N.A.  
RC STRAIN=HMI:IMSS;  
RX MEDLINE=21391859; PubMed=11500468;  
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,  
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,  
RA Tachibana H.;  
RA "Intermediate Subunit of the Gal/GalNac Lectin of Entamoeba  
RT histolytica Is a Member of a Gene Family Containing Multiple CXXC  
RT Sequence Motifs";  
RL Infect. Immun. 69:5892-5898 (2001).  
DR EMBL; AF337950; AAK92361.1; -.  
DR GO; GO:0005229; F:sugar binding; IEA.  
DR GO; GO:0007157; P:heterophilic cell adhesion; IEA.  
DR InterPro: IPR006209; EGF like.  
DR InterPro: IPR009030; GroW\_fac\_recep.  
DR PROSITE: PS01186; EGF\_2; 1.  
KW Lectin.  
SQ SEQUENCE 1101 AA; 119512 MW; C8B6F5CDBE56AEC CRC64;  
  
Query Match 11.7%; Score 177; DB 5; Length 1101;  
Best Local Similarity 26.6%; Pred. No. 1.2e-07;  
Matches 55; Conservative 22; Mismatches 80; Indels 50; Gaps 10;

```
QY 44 GGATCSD---YNGCL---SC---KPLRPFALB---RIGMKQIGVCLSSCPGYYGTR 89
Db 773 GTCCSCDLSKYPCCKKTDSCNVDSTRGFIYATCSDGFSRSPSYNCTTCTKSNYI-PK 831
QY 90 YPDINKTKCKACDCTCFNKNFCTK-----CKSGFYHLHLGKCLDNCPEGLEA 136
Db 832 EGEKNGCAKCDKCATCSDKDTCLTCAADPLKVGSKDCKCTGYTMSNGBC-----KPC 884
QY 137 NNHTMECVSIVHCEVSEWNPWSPCTKKG-KTC--GPKRGTTETRVREIIQHPSAKGNLCPP 193
Db 885 TNHCSECSAABCTVCSYDYKVISGNCNSVDGF-----YFDEINGTCIPC 932
QY 194 TNETRKTCTVORCKCKGKGRKKGRK 220
Db 933 TSPTCKCVGVKDCQEQTGCNSEKK 959
```

## RESULT 15

```
Q8SZS2 PRELIMINARY; PRT; 1376 AA.
AC Q8SZS2
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE LD30182p.
GN FUR2 OR CG4235 OR CG18734.
OS Drosophila melanogaster (Fruit fly).
OC Rukaryota; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phoumenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
DR EMBL; AY070553; AAL48024.1; -.
DR FlyBase; FBgn0004598; Fur2.
DR GO; GO:0004276; F: furin activity; IDA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR006212; Furin_repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF01483; P_protein; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM0261; FU_10.
DR PROSITE; PS00190; CYTOCHROME_C; 3.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine_protease.
SQ SEQUENCE 1376 AA; 149716 MW; B6704BA89A3A88FB CRC64;
```

```
Query Match 11.6%; Score 176.5; DB 5; Length 1376;
Best Local Similarity 28.0%; Pred. No. 1.7e-07;
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;
QY 3 HRLRLSWLFIILNPMYIGSNASRGRRQRMRHNPVSGCGCATCSDY-WGCLSCKPR 61
Db 734 HLHVID-LAVLCQCPDGYFENS-----RNRTCTVP-----CEPNCACQDHPHYCTSCDHH 783
QY 62 LPFALERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKADCDTCF--NKNPCTKCKSGF 119
```

```
Db 784 LVMHEHK-----CYSACPDLTYET---EDNKAFCCHSTCATCGPTDQDCITCRSSR 832
QY 120 YHLGLKCLDNCPEGLEBANNHTMECVSIVHCEVSEWNPWSPCTKKGKTCGPKRGTTETRV-R 178
Db 833 YAWQNKCLISCPDGFYADKKRLECM-----PCQBGCKTC-----TSNGVCS 873
QY 179 EIIQHPSAKGNLCPTTNETRKTCTVORKK-CQKGE 211
Db 874 ECLQWNT-----LNKRDKCIVSGSGGSESE 899
```

Search completed: June 29, 2004, 17:02:12  
Job time : 37.7876 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: June 29, 2004, 16:54:09 ; Search time 50.4502 Seconds  
(without alignments)  
1562.545 Million cell updates/sec

Title: US-09-894-912A-32  
Perfect score: 1535  
Sequence: 1 MHLRLISCFILNFMFYIG.....QQRRARDKQKSVSVSTVH 279

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04:.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1535	100.0	279	Aae13167	Aae13167 Mouse ste
2	1535	100.0	279	ABO44431	ABO44431 Mouse ste
3	1315.5	85.7	272	AAB99220	AAB99220 Human chr
4	1315.5	85.7	272	AAM78328	AAM78328 Human pro
5	1315.5	85.7	272	Aae13168	Aae13168 Human ste
6	1315.5	85.7	272	Aae13150	Aae13150 Human ste
7	1315.5	85.7	272	ABR62112	ABR62112 Human clo
8	1315.5	85.7	272	ABO44432	ABO44432 Human ste
9	1315.5	85.7	272	ABO44413	ABO44413 Human ste
10	1310.5	85.4	273	Aae13151	Aae13151 Human ste
11	1310.5	85.4	273	ABO44414	ABO44414 Human ste
12	1293	84.2	292	AAW85607	AAW85607 Secreted
13	1293	84.2	292	Aae13170	Aae13170 Human SCR
14	1293	84.2	292	ABP61846	ABP61846 Human pol
15	1293	84.2	292	ABR62114	ABR62114 Human sec
16	1293	84.2	292	ABO44434	ABO44434 Human SCR
17	1285	83.7	265	Aae13163	Aae13163 Human sec
18	1285	83.7	265	ABO44427	ABO44427 Human sec
19	1273	82.9	292	AAB93875	AAB93875 Human pro
20	1222.5	79.6	251	Aae13153	Aae13153 Human mat
21	1222.5	79.6	251	ABO44417	ABO44417 Human ste
22	1100	71.7	239	ABU52396	ABU52396 Human GPC
23	1045	68.1	195	ABU52398	ABU52398 Human GPC
24	1041	67.8	195	ABU52397	ABU52397 Human GPC
25	759	49.4	160	ABB11374	ABB11374 Human sec

26	759	49.4	160	4	AAM79312	AAM79312 Human pro
27	759	49.4	160	4	Aae13149	Aae13149 Human ste
28	759	49.4	160	6	ABO44415	ABO44415 Human ste
29	648	42.2	263	6	ABR62108	ABR62108 Secreted
30	648	42.2	263	6	ABR62115	ABR62115 Secreted
31	648	42.2	263	6	ABR58489	ABR58489 Human sec
32	648	42.2	263	7	ABO7919	ABO7919 Novel pro
33	640.5	41.7	243	6	ABR62110	ABR62110 Secreted
34	632	41.2	265	6	ABR62113	ABR62113 Mouse thr
35	629	41.0	229	4	Aae13162	Aae13162 Mouse thr
36	629	41.0	229	6	ABO44426	ABO44426 Human sec
37	571.5	37.2	243	5	Aae37115	Aae37115 Human sec
38	569.5	37.1	243	5	ABG76508	ABG76508 DNA encod
39	569.5	37.1	243	6	ABR62106	ABR62106 Secreted
40	565.5	36.8	243	6	Aae36166	Aae36166 Mouse SCR
41	565	36.8	250	6	ABR62101	ABR62101 Secreted
42	536	34.9	222	6	ABR62107	ABR62107 Secreted
43	531.5	34.6	229	6	ABR62102	ABR62102 Secreted
44	450	29.3	234	7	ADB76146	ADB76146 Novel hum
45	267	17.4	46	4	Aae13155	Aae13155 Human SCR

ALIGNMENTS

RESULT 1  
AAE13167  
ID AAE13167 standard; protein; 279 AA.  
XX AAE13167;

XX  
XX  
DT 28-JAN-2002 (first entry)  
XX  
DE Mouse stem cell growth factor-like protein.

XX  
KW Mouse; stem cell growth factor-like protein; antiinflammatory; nootropic;  
KW neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;  
KW vasotrophic; virucide; dermatological; tranquilliser; cerebroprotective;  
KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassemia;  
KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
KW supporting factor for the proliferation of stem cell.

XX  
OS Mus musculus.  
XX  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT Protein 22..279  
FT /label= Signal\_peptide  
FT /note= "Mouse mature stem cell growth factor-like protein"

WO200177169-A2.

18-OCT-2001.

05-APR-2001; 2001WO-US011208.

05-APR-2000; 2000US-00543774.

28-JUN-2000; 2000US-0215733P.

09-JAN-2001; 2001US-00757562.

05-FEB-2001; 2001US-0266614P.

(HYSE-) HYSEQ INC.

(KIRI ) KIRIN BEER KK.

Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;

PI	Stache-Crain B, Dickson M, Mize NK, Nishikawa M;	XX	Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
DR	WPI; 2001-657166/75.	XX	immunostimulant; vulnary; haematopoietic stem cell; gene therapy;
DR	N-PSDB; AAD21727.	XX	supporting factor for proliferation of stem cells; wound healing;
XX		XX	haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;
PT	Novel stem cell growth factor like polypeptides and polynucleotides for	XX	bone marrow transplantation; cord blood transplantation;
PT	identifying modulators useful for treating diseases such as Alzheimer's	XX	chronic granulomatous disease; duplicated immunodeficiency syndrome;
PT	disease, cancer, rheumatoid arthritis, osteoporosis.	XX	agammaglobulinemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
XX		XX	congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;
PS	Claim 28; Page 223-224; 232pp; English.	XX	epithelial cell growth; ovarian follicle development; nerve cell growth;
XX		XX	cartilage remodeling; bone growth; immunosuppression; mouse.
CC	The patent discloses novel stem cell growth factor-like proteins and	OS	Mus musculus.
CC	polynucleotides encoding them. Proteins of the invention are also known	XX	
CC	as supporting factor for the proliferation of stem cells (SCR-1). Stem	PH	Location/Qualifiers
CC	cell growth factor-like proteins are useful for supporting proliferation	FT	1. .21
CC	or survival of a stem cell or germ cell which is preferably primordial	FT	/label= Signal_peptide
CC	germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem	FT	22. .279
CC	cell, haematopoietic progenitor cell, pluripotent cell or totipotent	FT	/note= "Mature stem cell growth factor-like protein. This
CC	cell. The haematopoietic progenitor cell cultured using stem cell growth	FT	protein is specifically claimed in claim 9"
CC	factor-like proteins can replace as a graft for the bone marrow	XX	
CC	transplantation or cord blood transplantation for treating a variety of	PN	US2003044792-A1.
CC	diseases such as immunodeficiency syndrome, chronic granulomatous	XX	
CC	disease, duplicated immunodeficiency syndrome, agammaglobulinemia,	PD	06-MAR-2003.
CC	Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),	XX	
CC	thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia	XX	28-JUN-2001; 2001US-00894912.
CC	such as sickle cell anaemia, Gaucher's disease, lysosomal storage	XX	
CC	diseases such as mucopolysaccharidosis, adrenal white matter	XX	
CC	degeneration, a variety of cancer and tumours. Proteins of the invention	PR	28-JUN-2000; 2000US-0215733P.
CC	are useful for treating diseases such as Parkinson's disease, Alzheimer's	PR	05-FEB-2001; 2001US-0266614P.
CC	disease and other neurodegenerative diseases, thrombocytopaenia, immune	PR	05-APR-2001; 2001US-0282397P.
CC	deficiencies and disorders such as severe combined immunodeficiency	XX	
CC	(SCID) and autoimmune disorders such as multiple sclerosis, systemic	PA	(TANG/) TANG Y T.
CC	lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary	PA	(LABA/) LABAT I.
CC	inflammation. Sequences of the invention are also useful in gene therapy.	PA	(DRMA/) DRMANAC R T.
CC	The present sequence is stem cell growth factor-like protein from mouse	PA	(MIZE/) MIZE N.
XX		PA	(NISH/) NISHIKAWA M.
SQ	Sequence 279 AA;	PA	(CHAO/) CHAO C.
		XX	
	Query Match 100.0%; Score 1535; DB 4; Length 279;	PI	Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
	Best Local Similarity 100.0%; Pred. No. 6.2e-109;	XX	
	Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	
Qy	1 MHLLRLSCPIILNFMFYIGSQNASRRRRHPNVISQCGGCATCSYNGCLCKPR 60	XX	
Db	1 MHLLRLSCPIILNFMFYIGSQNASRRRRHPNVISQCGGCATCSYNGCLCKPR 60	XX	
Qy	61 LFPVLRIGMKQIGVCLSCSPGYGYTRYPDINKCTCKVDCTCFNKNFCTKCKSGFYL 120	XX	
Db	61 LFPVLRIGMKQIGVCLSCSPGYGYTRYPDINKCTCKVDCTCFNKNFCTKCKSGFYL 120	XX	
Qy	121 HLGKCLDSCPEGLEANNHTMECVSIHVCEASESPWSPCKMKGTCTGFRGTGTRVRDIL 180	XX	
Db	121 HLGKCLDSCPEGLEANNHTMECVSIHVCEASESPWSPCKMKGTCTGFRGTGTRVRDIL 180	XX	
Qy	181 QHPSAKGNLCPPTSETRTCTIVORKKCSGEGKGRERKRRKKLKEERKETSSSSDSK 240	XX	
Db	181 QHPSAKGNLCPPTSETRTCTIVORKKCSGEGKGRERKRRKKLKEERKETSSSSDSK 240	XX	
Qy	241 GLESSITETPDQNKERQKKRARDKQKSVSVTVH 279	XX	
Db	241 GLESSITETPDQNKERQKKRARDKQKSVSVTVH 279	XX	
RESULT 2		XX	
ABO44431		XX	
ID	ABO44431 standard; protein; 279 AA.	XX	
XX		XX	
AC	ABO44431;	XX	
DT		XX	
DT	30-SEP-2003 (first entry)	XX	
XX		XX	
DE	Mouse stem cell growth factor-like protein.	XX	
XX		XX	

Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; mouse.

Mus musculus.

Key Peptide 1. .21 /label= Signal\_peptide

Protein 22. .279 /note= "Mature stem cell growth factor-like protein. This protein is specifically claimed in claim 9"

US2003044792-A1.

06-MAR-2003.

28-JUN-2001; 2001US-00894912.

28-JUN-2000; 2000US-0215733P.

05-FEB-2001; 2001US-0266614P.

05-APR-2001; 2001US-0282397P.

(TANG/) TANG Y T.

(LABA/) LABAT I.

(DRMA/) DRMANAC R T.

(MIZE/) MIZE N.

(NISH/) NISHIKAWA M.

(CHAO/) CHAO C.

Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

WPI; 2003-625403/59.

N-PSDB; ACH04327.

Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.

Claim 23; Page 80; 96pp; English.

The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aa from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically



CC engineered to express the SCR-1 polypeptide to support proliferation or  
 CC survival of a stem cell or germ cell and an implant comprising a cell  
 CC genetically engineered to express the SCR-1 polypeptide to support  
 CC proliferation or survival of a stem cell or germ cell. The SCR-1  
 CC polypeptide is useful for identifying a compound that binds to the SCR-1  
 CC polypeptide and for maintaining survival of or promoting proliferation of  
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic  
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound  
 CC healing. The human haematopoietic stem cell or human haematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of haematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia  
 CC due to enzyme defect, congenital anaemia such as sickle cell disease, Gaucher's  
 CC disease etc. The SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a mouse SCR-1  
 CC protein

XX SQ Sequence 279 AA;

Query Match 100.0%; Score 1535; DB 6; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-109; Indels 0; Gaps 0;  
 Matches 279; Conservative 0; Mismatches 0;  
 QY 1 MHLRLISCFPIILNFMFYIGSQNASRRRRRHPNVSQGGCGCATCDYNGCLSCKPR 60  
 DB 1 MHLRLISCFPIILNFMFYIGSQNASRRRRRHPNVSQGGCGCATCDYNGCLSCKPR 60  
 QY 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCDTCFNKNFCTCKSGFYL 120  
 DB 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCDTCFNKNFCTCKSGFYL 120  
 QY 121 HLKCLDSCPEGLEANNHTMECVSIHCEASESWSPQMKGKTCGPKGTETVRDIL 180  
 DB 121 HLKCLDSCPEGLEANNHTMECVSIHCEASESWSPQMKGKTCGPKGTETVRDIL 180  
 QY 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGERKGRERKRLKLNKEERKETSSSDSK 240  
 DB 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGERKGRERKRLKLNKEERKETSSSDSK 240  
 QY 241 GLESSIETPDQENKERRQOQKRRARDKQKQSVSVTVH 279  
 DB 241 GLESSIETPDQENKERRQOQKRRARDKQKQSVSVTVH 279

RESULT 3  
 AAB99220  
 ID AAB99220 standard; protein; 272 AA.

XX AC AAB99220;  
 XX DT 07-SEP-2001 (first entry)  
 XX DE Human thrombospondin-30.  
 XX KW Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;  
 XX KW malignant tumour; haemopathy; HIV infection; immunological disease;  
 XX KW inflammation disease.  
 XX OS Homo sapiens.  
 XX PN WO200140294-A1.  
 XX PD 07-JUN-2001.  
 XX PF 27-NOV-2000; 2000WO-CN000494.  
 XX PR 29-NOV-1999; 99CN-00124148.

XX (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2001-397948/42.

XX N-PSDB; AAH45131.

XX Human thrombospondin-30 and polynucleotide is useful in diagnosis and  
 XX treatment of, e.g., malignant tumor, hemopathy, HIV infection,  
 XX immunological diseases and various inflammatory diseases.

XX Claim 1; Page 27-28; 33pp; Chinese.

XX The present sequence is the protein sequence for human thrombospondin-30.  
 XX Thrombospondin-30 protein and coding sequence are useful in the diagnosis  
 XX and treatment of malignant tumor, haemopathy, HIV infection,  
 XX immunological diseases and various inflammation diseases. In addition,  
 XX thrombospondin-30 protein may be used for screening mimics, agonists,  
 XX antagonists or inhibitors, or for use in peptide fingerprinting.  
 XX identification. The thrombospondin-30 coding sequence may be used as  
 XX primers for nucleic acid amplification reaction or as probes for  
 XX hybridisation reaction, or in producing gene chips or microarrays

XX SQ Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;  
 Best Local Similarity 87.1%; Pred. No. 3.1e-92;  
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;  
 QY 1 MHLRLISCFPIILNFMFYIGSQNASRRRRRHPNVSQGGCGCATCDYNGCLSCKPR 60  
 DB 1 MHLRLISCFPIILNFMFYIGSQNASRRRRRHPNVSQGGCGCATCDYNGCLSCKPR 60  
 QY 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCDTCFNKNFCTCKSGFYL 120  
 DB 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCDTCFNKNFCTCKSGFYL 120  
 QY 121 HLKCLDSCPEGLEANNHTMECVSIHCEASESWSPQMKGKTCGPKGTETVRDIL 180  
 DB 121 HLKCLDSCPEGLEANNHTMECVSIHCEASESWSPQMKGKTCGPKGTETVRDIL 180  
 QY 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGERKGRERKRLKLNKEERKETSSSDSK 240  
 DB 181 QHPSA--KGNLCPTTNETRTCTVQKCKSGERKGRERKRLKLNKEERKETS--AIPDSK 236  
 QY 241 GLESSIETPDQENKERRQOQKRRARDKQKQSVSVTVH 279  
 DB 237 SLESSKEIPQENK--QQQKRRVQDK-QKSVSVTVH 272

RESULT 4  
 AAM78328  
 ID AAM78328 standard; protein; 272 AA.

XX AC AAM78328;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human protein SEQ ID NO 990.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK51461.  
 XX  
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 XX Claim 20; Page 3214-3215; 6221pp; English.  
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3656 (AAW80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 SQ Sequence 272 AA;  
 Query Match 85.7%; Score 1315.5; DB 4; Length 272;  
 Best Local Similarity 87.1%; Pred. NO. 3.1e-92;  
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;  
 QY 1 MHLRLISCFITILFMFYTGSQNASGRQRORMHFNVSQCGGCATCSDYNGCLSKR 60  
 DB 1 MHLRLISMLFITILFMFYTGSQNASGRQRORMHFNVSQCGGCATCSDYNGCLSKR 60  
 QY 61 LPFVLRIQKQIGVCLSSCPGYGYTRYPDINKCTKCKVDCTCFNKNFCTKCKSGPYL 120  
 DB 61 LPFALLERIGKQIGVCLSSCPGYGYTRYPDINKCTKCKADCTCFNKNFCTKCKSGPYL 120  
 QY 121 HLGKCLDSCPEGLANNHTMBCVSIHVCRASNSPWSPCMKGKTCGPKRGTTETRVRI 180  
 DB 121 HLGKCLDNCPEGLANNHTMBCVSIHVCEVSENNPWSPTCKGTCGPKRGTTETRVREII 180  
 QY 181 QHPSAKGNLCPPYSETCTIVQKCKSGERGKGRERKELKKEKETSSSSDSK 240  
 DB 181 QHPSA--KGNLCPPYETTRKCTVQRKCKGGERGKGRERKRPKNKGSKE--AIPDSK 236  
 QY 241 GLESSLETDPQENKRCQQQKRRARDKQKSVSVSTVH 279  
 DB 237 SLESSKEIPEQRNK--QQQKREKQDK-QKSVSVSTVH 272  
 RESULT 5  
 ID AAB13168  
 XX AAB13168 standard; protein; 272 AA.  
 AC AAE13168;  
 XX  
 XX 28-JAN-2002 (first entry)  
 DT  
 XX

DE Human stem cell growth factor-like protein #4.  
 XX Human; stem cell growth factor-like protein; antiinflammatory; neutrophic;  
 KW neuroprotective; vulnerary; cytosstatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 XX supporting factor for the proliferation of stem cell.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 PT Peptide 1..21  
 PT Protein /label= Signal\_peptide  
 PT 22..272  
 FT /note= "Human mature stem cell growth factor-like  
 FT protein"  
 XX WO200177169-A2.  
 XX 18-OCT-2001.  
 XX 05-APR-2001; 2001WO-US011208.  
 XX 05-APR-2000; 2000US-00543774.  
 PR 28-JUN-2000; 2000US-0215733P.  
 PR 09-JAN-2001; 2001US-00757562.  
 PR 05-FEB-2001; 2001US-0266614P.  
 XX (HYSE-) HYSEQ INC.  
 PA (KIRI ) KIRIN BEER KK.  
 XX  
 PI Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;  
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
 XX WPI; 2001-657166/75.  
 DR N-PSDB; AAD21728.  
 XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.  
 XX Claim 28; Page 226-227; 232pp; English.  
 XX The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumour. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary

CC inflammation. Sequences of the invention are also useful in gene therapy.  
CC The present sequence is stem cell growth factor-like protein from human  
XX  
SQ Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;  
Best Local Similarity 87.1%; Pred. No. 3.1e-92;  
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHURLISCFPIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60  
Db 1 MHURLISLWFLIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60  
QY 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGPYL 120  
Db 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGPYL 120  
QY 121 HLKCLDSCPEGLEANNHTMECVSIHVCEASESPWSPCKMKGKTCGFRGRTTRVRDIL 180  
Db 121 HLKCLDNCPEGLEANNHTMECVSIHVCEASESPWSPCTKKGTCGFRGRTTRVRDIL 180  
QY 181 QHPSAKGNLCPPTSETRTCTIVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240  
Db 181 QHPSA--KGNLCPPTWETRTCTVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240  
QY 241 GLESSIETPDQENKEROQOQKERRARDKQKQSVSVSTVH 279  
Db 237 SLESSKEIPEQRENK--QOQKGRKVQDK-QKQSVSVSTVH 272

RESULT 6  
AAE13150  
ID AAE13150 standard; protein; 272 AA.  
XX  
AC AAE13150;  
XX  
DT 28-JAN-2002 (first entry)  
XX  
DE Human stem cell growth factor-like protein #2.  
XX  
KW Human; stem cell growth factor-like protein; antiinflammatory; neutropic;  
KW neutropic; vulnary; cytostatic; anticonvulsant; immunostimulant;  
KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;  
KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
KW supporting factor for the proliferation of stem cell.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21 /label= Signal\_peptide  
FT Protein 22..272 /note= "Human mature stem cell growth factor-like protein"

WO200177169-A2.  
18-OCT-2001.  
05-APR-2001; 2001WO-US011208.  
05-APR-2000; 2000US-00543774.  
28-JUN-2000; 2000US-0215733P.  
09-JAN-2001; 2001US-00757562.  
05-FEB-2001; 2001US-0266614P.

XX  
PA (HYSB-) HYSQ INC.  
XX (KIRI ) KIRIN BEER KK.  
XX  
PI Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;  
PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
XX WPI; 2001-657166/75.  
DR N-PSDB; AAD21724.  
XX  
XX Novel stem cell growth factor like polypeptides and polynucleotides for  
PT identifying modulators useful for treating diseases such as Alzheimer's  
PT disease, cancer, rheumatoid arthritis, osteoporosis.  
XX  
PS Claim 28; Page 211-212; 232pp; English.  
XX  
CC The patent discloses novel stem cell growth factor-like proteins and  
CC polynucleotides encoding them. Proteins of the invention are also known  
CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
CC cell growth factor-like proteins are useful for supporting proliferation  
CC or survival of a stem cell or germ cell which is preferably primordial  
CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
CC factor-like proteins can replace as a graft for the bone marrow  
CC transplantation or cord blood transplantation for treating a variety of  
CC diseases such as immunodeficiency syndrome, chronic granulomatous  
CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
CC diseases such as mucopolysaccharidosis, adrenal white matter  
CC degeneration, a variety of cancer and tumours. Proteins of the invention  
CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, thrombocytopaenia, immune  
CC deficiencies and disorders such as severe combined immunodeficiency  
CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
CC inflammation. Sequences of the invention are also useful in gene therapy.  
CC The present sequence is stem cell growth factor-like protein from human  
XX  
SQ Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;  
Best Local Similarity 87.1%; Pred. No. 3.1e-92;  
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHURLISCFPIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60  
Db 1 MHURLISLWFLIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60  
QY 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGPYL 120  
Db 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGPYL 120  
QY 121 HLKCLDSCPEGLEANNHTMECVSIHVCEASESPWSPCKMKGKTCGFRGRTTRVRDIL 180  
Db 121 HLKCLDNCPEGLEANNHTMECVSIHVCEASESPWSPCTKKGTCGFRGRTTRVRDIL 180  
QY 181 QHPSAKGNLCPPTSETRTCTIVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240  
Db 181 QHPSA--KGNLCPPTWETRTCTVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240  
QY 241 GLESSIETPDQENKEROQOQKERRARDKQKQSVSVSTVH 279  
Db 237 SLESSKEIPEQRENK--QOQKGRKVQDK-QKQSVSVSTVH 272

RESULT 7  
ABR62112  
ID ABR62112 standard; protein; 272 AA.  
XX  
AC ABR62112;

XX 18-AUG-2003 (first entry)  
 XX Human clone 1 thrombospondin protein #23.  
 XX Human; secreted; stem cell growth factor; cytostatic; haemostatic; neuroprotective; immunostimulant; leukaemia; haemophilia; cancer; degenerative disease; Alzheimer's disease; food supplement; immunological disorder; thrombospondin.  
 OS Homo sapiens.  
 XX WO2003029405-A2.  
 XX 10-APR-2003.  
 XX 30-AUG-2002; 2002WO-US027746.  
 XX 30-AUG-2001; 2001US-0316368P.  
 PR 10-DEC-2001; 2001US-0339739P.  
 PR 19-APR-2002; 2002US-00125852.  
 XX (HYSB-) HYSEQ INC.  
 PA Tang YT;  
 PI WPI; 2003-381616/36.  
 XX New stem cell growth factor-like polypeptides and polynucleotides, useful for treating e.g. leukemia, hemophilia and degenerative diseases like Alzheimer's disease, and for inducing immune response.  
 XX Disclosure; Fig 1; 15pp; English.  
 XX The invention relates to new stem cell growth factor-like polypeptides and polynucleotides. The stem cell growth factor-like polypeptides and polynucleotides are useful for inducing differentiation of embryonic and adult stem cells to give rise to different cell types, for treating e.g. leukaemia, haemophilia and degenerative diseases like Alzheimer's disease. They are also useful for generating new tissues and organs that may aid patients in need of transplanted tissues. The polynucleotides are useful as hybridisation probes, oligomers or primers for PCR, for chromosome and gene mapping, in recombinantly producing protein, in generating antisense DNA or RNA, in diagnostics as expressed sequence tags for identifying expressed genes, and for inducing immune response. The polypeptides are useful for generating antibodies that specifically bind the polypeptide, as molecular weight markers, and as a food supplement (e.g. protein or amino acid supplement, and as a carbon, nitrogen or carbohydrate source). Compositions comprising the polypeptides or polynucleotides are useful for the diagnosis, treatment or prevention of cancers, and other immunological disorders. The current sequence represents a human clone 1 thrombospondin protein

XX SQ Sequence 272 AA;  
 Query Match 85.7%; Score 1315.5; DB 6; Length 272;  
 Best Local Similarity 87.1%; Pred. No. 3.1e-92;  
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMFYIGSNASRGRRQRMHPNVISQGGCATCSYNGCLSKPR 60  
 DB 1 MHLRLISWLFILNFMFYIGSNASRGRRQRMHPNVISQGGCATCSYNGCLSKPR 60  
 QY 61 LFFVLERIGHKQIGVCLSSCPGSGYGYTRYDINKTKCTVDCTCFNKNFCTKCSGYL 120  
 DB 61 LFFALERIGHKQIGVCLSSCPGSGYGYTRYDINKTKCTVDCTCFNKNFCTKCSGYL 120  
 QY 121 HLGKCLDSCPEGLAANNHMECVSIHVCEASEMSFSPCKMKGKTCGFKGTETRVADIL 180  
 DB 121 HLGKCLDNCPEGLAANNHMECVSIHVCESEWNPSPCTCKGKTCGFKGTETRVREII 180  
 QY 181 QHPSAKGKGLCPPTSTRCTIVORKFCKSGERKGRKGRKGLNKERKFTSSSDSK 240

DB 181 QHPSA--KGNLCPTNETRKCTVQRKKCKQGRGKKGRERKRRKPKNGSKB--AIPDSK 236  
 QY 241 GLESSIETPDQQENKERQQQKRRARDKQKSVSVSTVH 279  
 DB 237 SLESSKEIPEQRNK--QQQKRRKVQDK-QKSVSVSTVH 272

RESULT 8  
 ID ABO44432 standard; protein; 272 AA.  
 XX ABO44432;  
 XX 30-SEP-2003 (first entry)  
 XX Human stem cell growth factor-like protein, SCR 1 #4.  
 XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia; bone marrow transplantation; cord blood transplantation;  
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; human.  
 OS Homo sapiens.  
 XX Key  
 FH Peptide 1..21 Location/Qualifiers  
 FT /label= signal\_peptide  
 FT Protein 22..272 /note= "Mature stem cell growth factor-like protein. This protein is specifically claimed in claim 9"

US2003044792-A1.  
 06-MAR-2003.  
 28-JUN-2001; 2001US-00894912.  
 28-JUN-2000; 2000US-0215733P.  
 05-FEB-2001; 2001US-0266614P.  
 05-APR-2001; 2001US-0282397P.  
 (TANG/) TANG Y T.  
 (LABA/) LABAT I.  
 (DRMA/) DRMANAC R T.  
 (MIZE/) MIZE N.  
 (NISH/) NISHIKAWA M.  
 (CHAO/) CHAO C.  
 Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;  
 WPI; 2003-625403/59.  
 N-PSDB; ACH04328.  
 Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.  
 Claim 23; Page 82; 96pp; English.  
 The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the

complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aas from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell and an implant comprising a cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell. The SCR-1 polypeptide is useful for identifying a compound that binds to the SCR-1 polypeptide and for maintaining survival of or promoting proliferation of a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic progenitor cell. The SCR-1 polypeptide is useful for promoting wound healing. The human haematopoietic stem cell or human haematopoietic progenitor cell culture using the SCR-1 polypeptide can replace as a graft for the conventional bone marrow transplantation or cord blood transplantation. The transplantation of haematopoietic stem cells can be employed as a therapy for treating diseases such as chronic granulomatous diseases, duplicated immunodeficiency syndrome, agammaglobulinemia, Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia such as sickle cell disease, Gaucher's disease etc. The SCR-1 polypeptide is useful for cell growth and morphogenesis, including tissue specific stem cell growth, epithelial cell growth and regulation, ovarian follicle development, promoting nerve cell growth, sustaining neuronal populations, cartilage remodeling, bone growth and immunosuppression. The present sequence is a Human SCR-1 protein.

Sequence 272 AA:

Query Match	85.7%;	Score	1315.5;	DB	6;	Length	272;
Best Local Similarity	87.1%;	Pred. No.	3.1e-92;				
Matches	243;	Conservative	11;	Mismatches	18;	Indels	7;
						Gaps	4;
Qy	1	MHLRLISCFPIILFMFVIGSQNASRGRQRORRMHPNVSQGGGCATCSDYNGCLSCKPR	60				
Db	1	MHLRLISWLFILFMFVIGSQNASRGRQRORRMHPNVSQGGGCATCSDYNGCLSCKPR	60				
Qy	61	LPFVLERIGMKQIGVCLSSPSGYYGTRYPDINCKTKCKVDCDTCFNKNFCTCKCKSGPYL	120				
Db	61	LPFALERIGMKQIGVCLSSPSGYYGTRYPDINCKTKCKADCDTCFNKNFCTCKCKSGPYL	120				
Qy	121	HLGKLDSCPGLKANNHTECVSIVHCEASGWSWSPCKMKKTCGPKRGTTETVRDIL	180				
Db	121	HLGKLDNCPGLKANNHTECVSIVHCEVSEWNPSPCTKKGKTCGPKRGTTETVRDII	180				
Qy	181	QHPSAKGKGNLCPTTSETRTICIVORKKCKSGRKGKGRKRKRLKLNKEBKETSSSSSS	240				
Db	181	QHPSA--KGNLCPTTNETKTIVQRKCKQKGRGKGRKRKRPKNKESKE--AIPDSK	236				
Qy	241	GLESSITPDQENKRRQQQKRRARDKQKSVSVSTVH	279				
Db	237	SLRSSKRIPIRENNK--COOKRRKVODK--OKSVSVSTVH	272				

## RESULT 9

ABO44413

ID ABO44413 standard; protein; 272 AA.

XX

AC ABO4413;

30-SEP-2003 (first entry)

Human stem cell growth factor-like protein, SCR 1 #1.

Stem cell growth factor-like protein; anitranemic; anti-HIV; SCF-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklecella; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; human

**Homo sapiens.**

US2003044792-A1.

06-MAR-2003

28-JUN-2001: 2001US-00894912

28-JUN-2000: 2000TS-0215733P

28-JUN-2000; 2000US-0215153E;  
05-FEB-2001; 2001US-0266614P;

05-APR-2001; 2001US-0282397P.

(TANG/) TANG Y T.

(LABA/) LABAT I.

(DRMA/) DRMANAC R T.

(MIZB/) MIZB N.

(NISH/) NISHIKAWA M.

(CHAO//) CHAO C.

Tang VT, Labat T, Dymanac RT, Mize N, Nishikawa M, Chao C:

WPI: 2003-625403/59

WFI; 2003-023403/33.  
N-PSDB; ACH04323, ACH04324.

Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.

Claim 23: Fig 3: 96pp: English.

The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aas from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO444328 and ABO444429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell and an implant comprising a cell genetically engineered to express the SCR-1 polypeptide to support

CC proliferation or survival of a stem cell or germ cell. The SCR-1  
 CC polypeptide is useful for identifying a compound that binds to the SCR-1  
 CC polypeptide and for maintaining survival of or promoting proliferation of  
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic  
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound  
 CC healing. The human haematopoietic stem cell or human haematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of haematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia,  
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's  
 CC disease etc. The SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a Human SCR-1  
 CC protein  
 XX  
 SQ Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 6; Length 272;  
 Best Local Similarity 87.1%; Pred. No. 3.1e-92;  
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHRLISCPFIILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKR 60  
 DB 1 MHRLISWLFILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKR 60  
 QY 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKCKVDCDTFKNKFCCKSGFYL 120  
 DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKCKVDCDTFKNKFCCKSGFYL 120  
 QY 121 HLGKCLDSCPEGLNHNHMECVSIHVCEASEWSPSPCKMKGKTCGFKGTETVRDIL 180  
 DB 121 HLGKCLDSCPEGLNHNHMECVSIHVCEASEWSPSPCKMKGKTCGFKGTETVRDIL 180  
 QY 181 QHPSAKGKGNLCPTSETRTTCIVORKKSKGKRGKGRKRLKLNKBERKPTSSSSDSK 240  
 DB 181 QHPSA--KGNLCPTTETRTCTVQRKKQGERGKGRKRLKPNKGSKE--AIPDSK 236  
 QY 241 GLESSITPDQENKGRQOKRRARDKQKSVSVTVH 279  
 DB 237 SLESSKEIPEQRENK--QOKRKRVQDK-QKSVSVTVH 272

RESULT 10  
 AAE13151  
 ID AAE13151 standard; protein; 273 AA.

XX AAE13151;  
 AC AAE13151;  
 XX  
 DT 28-JAN-2002 (first entry)  
 XX  
 DE Human stem cell growth factor-like protein #3.

XX Human; stem cell growth factor-like protein; antiinflammatory; neutropic;  
 KW neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disease;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 KW supporting factor for the proliferation of stem cell.  
 OS Homo sapiens.

WO200171169-A2.

18-OCT-2001.

05-APR-2001; 2001WO-US011208.

05-APR-2000; 2000US-00543774.

28-JUN-2000; 2000US-0215733P.

09-JAN-2001; 2001US-00757562.

05-FEB-2001; 2001US-0266614P.

(HYSB-) HYSQ INC.

(KIRI ) KIRIN BEER KK.

Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;

Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

WPI; 2001-6571166/75.

N-PSDB; AAD21725.

Novel stem cell growth factor like polypeptides and polynucleotides for  
 identifying modulators useful for treating diseases such as Alzheimer's  
 disease, cancer, rheumatoid arthritis, osteoporosis.

Claim 28; Page 214-215; 232pp; English.

The patent discloses novel stem cell growth factor-like proteins and  
 polynucleotides encoding them. Proteins of the invention are also known  
 as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 cell growth factor-like proteins are useful for supporting proliferation  
 or survival of a stem cell or germ cell which is preferably primordial  
 germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 cell. The haematopoietic progenitor cell cultured using stem cell growth  
 factor-like proteins can replace as a graft for the bone marrow  
 transplantation or cord blood transplantation for treating a variety of  
 diseases such as immunodeficiency syndrome, chronic granulomatous  
 disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 thalassemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 diseases such as mucopolysaccharidosis, adrenal white matter  
 degeneration, a variety of cancer and tumours. Proteins of the invention  
 are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 disease and other neurodegenerative diseases, thrombocytopaenia, immune  
 deficiencies and disorders such as severe combined immunodeficiency  
 (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 inflammation. Sequences of the invention are also useful in gene therapy.  
 The present sequence is stem cell growth factor-like protein from human

Sequence 273 AA;

Query Match 85.4%; Score 1310.5; DB 4; Length 273;  
 Best Local Similarity 87.1%; Pred. No. 7.5e-92;  
 Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 2 HRLISCPFIILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKRL 61  
 DB 3 HRLISWLFILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKRL 62  
 QY 62 FFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKCKVDCDTFKNKFCCKSGFYL 121  
 DB 63 FFFALERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKCKVDCDTFKNKFCCKSGFYL 122  
 QY 122 LGKCLDSCPEGLNHNHMECVSIHVCEASEWSPSPCKMKGKTCGFKGTETVRDILQ 181  
 DB 123 LGKCLDSCPEGLNHNHMECVSIHVCEASEWSPSPCTKKGKTCGFKGTETVRDILQ 182  
 QY 182 HPSAKGKGNLCPTSETRTTCIVORKKSKGKRGKGRKRLKLNKBERKPTSSSSDSK 241  
 DB 183 HPSA--KGNLCPTTETRTCTVQRKKQGERGKGRKRLKPNKGSKE--AIPDSK 238



QY 242 LESSIETPDQENKEROOQKRRARDKQKSVSVSTVH 279  
 DB 239 LESSKEIPEQRENK--OOQKRRVQDK-QKSVSVSTVH 273

RESULT 11  
 ABO44414  
 ID ABO44414 standard; protein; 273 AA.  
 AC ABO44414;  
 XX 30-SEP-2003 (first entry)  
 DT Human stem cell growth factor-like protein, SCR 1 #2.  
 DE  
 XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
 KW immunostimulant; vulnary; haematopoietic stem cell; gene therapy;  
 KW supporting factor for proliferation of stem cells; wound healing;  
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;  
 KW bone marrow transplantation; cord blood transplantation;  
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;  
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;  
 KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;  
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;  
 KW cartilage remodeling; bone growth; immunosuppression; human.  
 XX Homo sapiens.  
 OS  
 XX US2003044792-A1.  
 XX 06-MAR-2003.  
 XX 28-JUN-2001; 2001US-00894912.  
 XX 28-FEB-2000; 2000US-0215733P.  
 PR 05-FEB-2001; 2001US-0266614P.  
 PR 05-APR-2001; 2001US-0282397P.  
 XX (TANG/) TANG Y T.  
 PA (LABA/) LABAT I.  
 PA (DRMA/) DRMANAC R T.  
 PA (MIZE/) MIZE N.  
 PA (NISH/) NISHIKAWA M.  
 PA (CHAO/) CHAO C.  
 XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;  
 PI WPI: 2003-625403/59.  
 XX N-PSDB; ACH04325.  
 XX Novel isolated polypeptide having stem cell growth factor activity,  
 PT useful for promoting wound healing, and as a medicine to proliferate or  
 FT support human hematopoietic stem cells or human hematopoietic progenitor  
 PT cells.  
 XX Claim 23; Page 72-73; 96pp; English.

The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44413), an

isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aas from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising of a stem cell or germ cell, an anti-SCR 1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell and an implant comprising a cell proliferation or survival of a stem cell or germ cell. The SCR-1 polypeptide is useful for identifying a compound that binds to the SCR-1 polypeptide and for maintaining survival of or promoting proliferation of a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic progenitor cell. The SCR-1 polypeptide is useful for promoting wound healing. The human haematopoietic stem cell or human haematopoietic progenitor cell culture using the SCR-1 polypeptide can replace as a graft for the conventional bone marrow transplantation or cord blood transplantation. The transplantation of haematopoietic stem cells can be employed as a therapy for treating diseases such as chronic granulomatous diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia, Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's disease etc. the SCR-1 polypeptide is useful for cell growth and morphogenesis, including tissue specific stem cell growth, epithelial cell growth and regulation, ovarian follicle development, promoting nerve cell growth, sustaining neuronal populations, cartilage remodeling, bone cell growth and immunosuppression. The present sequence is a human SCR-1 protein

Query Match 85.4%; Score 1310.5; DB 6; Length 273;  
 Best Local Similarity 87.1%; Pred. No. 7.5e-92;  
 Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;  
 Sequence 273 AA;  
 QY 2 HRLISCFPIILNFMETIGSQNASRGRQRORHPNVNVSQCGGCATCSYNGCLSKCPRL 61  
 DB 3 HRLISWLFILNFMETIGSQNASRGRQRORHPNVNVSQCGGCATCSYNGCLSKCPRL 62  
 QY 62 PFVLERIGMKQIGVCLSCPSGYTGTRYPDINKCTKCKVDCDTCFNKNPCTKCKSGFYLH 121  
 DB 63 PFALERIGMKQIGVCLSCPSGYTGTRYPDINKCTKCKADCDTCFNKNPCTKCKSGFYLH 122  
 QY 122 LGKCLDSCPEGLEANNHTMECVSIHVCEASBWSWSPSCMKKGTGCPKRGTRVRDILQ 181  
 DB 123 LGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFKRGTRVREIIQ 182  
 QY 182 HPSAKGKYLCPPTSETTCTVORCKSKGRGKGRKRLKKEKKEKSSSSDSKG 241  
 DB 183 HPSA--KGNLCPTNTRKCTVQRKKCQKGRGKGRKRLKKNKSGSK--AIPDSKS 238  
 QY 242 LESSIETPDQENKEROOQKRRARDKQKSVSVSTVH 279  
 DB 239 LESSKEIPEQRENK--OOQKRRVQDK-QKSVSVSTVH 273

RESULT 12  
 AAW85607  
 ID AAW85607 standard; protein; 292 AA.  
 AC AAW85607;  
 XX 02-MAR-1999 (first entry)  
 DT Secreted protein clone da238\_6.  
 DE  
 KW Clone; secreted protein; protein factor; cytokine; lymphokine;  
 KW interferon; colony stimulating factor; CSF; interleukin; cloning;  
 KW tumour invasion; tumour suppression; immune boosting.  
 XX Homo sapiens.  
 OS





Db 1 MHLRLISWLFILMFYIGSNASRGRRRRMHPNVSGQGCGCATCDYNGCLSCKPR 60  
Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYPIPKCTCKVDCDTCFKNKFTCKKSGPYL 120  
Db 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYPIPKCTCKVDCDTCFKNKFTCKKSGPYL 120  
Qy 121 HLKCLDSCPEGLNNTMBCVSIHVCEASEWSPWSPCMKKGKTKGKGTGFRGTETVRDIL 180  
Db 121 HLKCLDSCPEGLNNTMBCVSIHVCEASEWSPWSPCMKKGKTKGKGTGFRGTETVRDIL 180  
Qy 181 QHPSAKGKGLCPPTSETTRTCIVQKCKSGRGGKGRKRRKRLKLNKEERKETSSSSDSK 240  
Db 181 QHPSA--KGNLCPTTNETKCTVQKCKSGRGGKGRKRRKRLKLNKEERKETSSSSDSK 240  
Qy 241 GLESSIETPDQENKQKQKRRARDKQKSVSVS 276  
Db 237 SLESSKIPEQRENK--QKQKRVQDKQKSGIYVT 270

## RESULT 14

ABP61846

ID ABP61846 standard; protein; 292 AA.

XX

AC ABP61846;

XX

DT 04-OCT-2002 (first entry)

XX

DE Human polypeptide SEQ ID NO 200.

XX

Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;  
antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;  
neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;  
antitumor; fungicide; antidiabetic; antiaesthetic; antiallergic;  
immunostimulant; antiparasitic; secreted protein; transmembrane protein;  
cytokine; cell proliferation; cell differentiation; autoimmune disease;  
stem cell; growth factor; nervous system disease; neuropathy;  
Alzheimer's disease; Parkinson's disease; Huntington's disease;  
osteoporosis; severe combined immunodeficiency; SCID; infection;  
multiple sclerosis; rheumatoid arthritis; gene therapy.

OS Homo sapiens.

XX US2002065394-A1.

XX

XX 30-MAY-2002.

XX

XX 22-DEC-2000; 2000US-00745763.

XX

XX 18-MAR-1998; 98US-00040963.

XX

XX (JACO/) JACOBS K.

XX (MCCO/) MCCOY J M.

XX (LAVA/) LAVALLIE E R.

XX (COLL/) COLLINS-RACIE L A.

XX (EVAN/) EVANS C.

XX (MERB/) MERBERG D.

XX (TREAC/) TREACY M.

XX (SPAUL/) SPAULDING V.

XX

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M, Spaulding V;

XX

XX WPI: 2002-582343/62.

XX N-PSDB; ABQ92060.

XX

Novel secreted or transmembrane protein and polynucleotide encoding the  
protein, useful for diagnosis and treatment of neurological disorders,  
cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.

XX

XX Claim 207; Page 203-204; 284pp; English.

XX

The invention relates to human secreted or transmembrane protein (I),  
their fragments and is encoded by specific complementary deoxyribonucleic

CC acid (cDNA) inserts (II), where the protein is substantially free from  
CC other mammalian proteins. (I) are useful for preventing, treating or  
CC ameliorating a medical condition, especially immunological treatment or  
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,  
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,  
CC stem cell growth factor activity and activin or inhibin-related  
CC activities. (I) can be used to manipulate stem cells in culture to give  
CC rise to neuroepithelial cells that can be used to augment or replace  
CC cells damaged by illness, autoimmune disease, accidental damage or  
CC genetic disorders. (I) induces the proliferation of neural cells and  
CC regeneration of nerve and brain tissue and is useful for the treatment of  
CC central and peripheral nervous system diseases and neuropathies, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis and is useful for treating myeloid  
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia  
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,  
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or  
CC periodontal disease. (I) is also useful for gut protection or  
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury  
CC in various tissues, various immune deficiencies and disorders including  
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,  
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,  
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,  
CC such as asthma or other respiratory problems. (II) is useful to express  
CC recombinant protein, as markers for tissues in which the corresponding  
CC protein is preferentially expressed and in gene therapy. The present  
CC sequence is that of a polypeptide of the invention  
XX

SQ Sequence 292 AA;

Query Match 84.2%; Score 1293; DB 5; Length 292;

Best Local Similarity 85.1%; Pred No. 1.7e-90;

Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;

Qy 1 MHLRLISWLFILMFYIGSNASRGRRRRMHPNVSGQGCGCATCDYNGCLSCKPR 60

Db 1 MHLRLISWLFILMFYIGSNASRGRRRRMHPNVSGQGCGCATCDYNGCLSCKPR 60

Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYPIPKCTCKVDCDTCFKNKFTCKKSGPYL 120

Db 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYPIPKCTCKVDCDTCFKNKFTCKKSGPYL 120

Qy 121 HLKCLDSCPEGLNNTMBCVSIHVCEASEWSPWSPCMKKGKTKGKGTGFRGTETVRDIL 180

Db 121 HLKCLDSCPEGLNNTMBCVSIHVCEASEWSPWSPCMKKGKTKGKGTGFRGTETVRDIL 180

Qy 181 QHPSAKGKGLCPPTSETTRTCIVQKCKSGRGGKGRKRRKRLKLNKEERKETSSSSDSK 240

Db 181 QHPSA--KGNLCPTTNETKCTVQKCKSGRGGKGRKRRKRLKLNKEERKETSSSSDSK 240

Qy 241 GLESSIETPDQENKQKQKRRARDKQKSVSVS 276

Db 237 SLESSKIPEQRENK--QKQKRVQDKQKSGIYVT 270

## RESULT 15

ABR62114

ID ABR62114 standard; protein; 292 AA.

XX

AC ABR62114;

XX

DT 18-AUG-2003 (first entry)

XX

XX Human secreted protein clone da\_288\_6 #25.

XX Human; secreted; stem cell growth factor; cytostatic; haemostatic;

XX neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;

XX degenerative disease; Alzheimer's disease; food supplement;

XX immunological disorder.

XX

OS Homo sapiens.

```
XX WO2003029405-A2.
XX
XX
XX PD 10-APR-2003.
XX
XX PF 30-AUG-2002; 2002WO-US027746.
XX
XX PR 30-AUG-2001; 2001US-0316368P.
XX
XX PR 10-DEC-2001; 2001US-0339739P.
XX
XX PR 19-APR-2002; 2002US-00125852.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT;
XX
XX WPI; 2003-381616/36.
XX
XX New stem cell growth factor-like polypeptides and polynucleotides, useful
PT for treating e.g. leukemia, hemophilia and degenerative diseases like
PT Alzheimer's disease, and for inducing immune response.
XX
XX Disclosure; Fig 2; 151pp; English.
XX
XX The invention relates to new stem cell growth factor-like polypeptides
CC and polynucleotides. The stem cell growth factor-like polypeptides and
CC polynucleotides are useful for inducing differentiation of embryonic and
CC adult stem cells to give rise to different cell types, for treating e.g.
CC leukaemia, haemophilia and degenerative diseases like Alzheimer's
CC disease. They are also useful for generating new tissues and organs that
CC may aid patients in need of transplanted tissues. The polynucleotides are
CC useful as hybridisation probes, oligomers or primers for PCR, for
CC chromosome and gene mapping, in recombinantly producing protein, in
CC generating antisense DNA or RNA, in diagnostics as expressed sequence
CC tags for identifying expressed genes, and for inducing immune response.
CC The polypeptides are useful for generating antibodies that specifically
CC bind the polypeptide, as molecular weight markers, and as a food
CC supplement (e.g. protein or amino acid supplement, and as a carbon,
CC nitrogen or carbohydrate source). Compositions comprising the
CC polypeptides or polynucleotides are useful for the diagnosis, treatment
CC or prevention of cancers, and other immunological disorders. The current
CC sequence represents a human secreted protein clone da_288_6
XX
XX Sequence 292 AA;
Query Match 84.2%; Score 1293; DB 6; Length 292;
Best Local Similarity 85.1%; Pred. No. 1.7e-90;
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;
QY 1 MHLRLISCFPIILFMFMEYIGSQNASRGRQRMRHFNVSQCGGCATCSDYNGCLSCKPR 60
DB 1 MHLRLISWLFPIILFMFMEYIGSQNASRGRQRMRHFNVSQCGGCATCSDYNGCLSCKPR 60
QY 61 LFPVLERIGMKQIGVCLSCPSSGYGTRYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYL 120
DB 61 LFPALERIGMKQIGVCLSCPSSGYGTRYPDINKCTKCKADCDTCFNKNFCTKCKSGPYL 120
QY 121 HLGKCLDSCPEGLHANNHTMECVSIHVCEASEWSFWSCKKGGKTCGPKRGTTETVRDIL 180
DB 121 HLGKCLDNCPEGLHANNHTMECVSIHVCESENNPWSPTCKGKTCGPKRGTTETVRREII 180
QY 181 QHPSARGKNLCPTTSETRTCTIVQRKCKSGRGRKGRKRLKLNKEKERTSSSSDSK 240
DB 181 QHPSA--KGNLCPTTETRTCTIVQRKCKGKGRKGRKRLKLNKEKERTSSSSDSK 240
QY 241 GLESSITPTDQENKERQOQKRRARDKQKQSVSVS 276
DB 237 SLESSKEIPEQRENK--QOQKRRKVQDKQKSGIEVT 270
```

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	179	11.7	969	2	US-08-284-941-2		Sequence 2, Appli
2	179	11.7	969	2	US-08-447-642-2		Sequence 2, Appli
3	179	11.7	969	3	US-09-236-503-2		Sequence 2, Appli
4	179	11.7	969	5	PCF-US93-02147A-2		Sequence 2, Appli
5	155	10.1	288	1	US-08-368-852-15		Sequence 15, Appli
6	154	10.0	799	2	US-08-525-940-23		Sequence 23, Appli
7	154	10.0	799	2	US-08-976-838-23		Sequence 23, Appli
8	154	10.0	881	2	US-08-525-940-21		Sequence 21, Appli
9	154	10.0	881	2	US-08-976-838-21		Sequence 21, Appli
10	154	10.0	915	2	US-08-525-940-18		Sequence 18, Appli
11	154	10.0	915	2	US-08-976-838-18		Sequence 18, Appli
12	154	10.0	915	4	US-09-214-5558-2		Sequence 2, Appli
13	154	10.0	915	4	US-09-214-5558-7		Sequence 7, Appli
14	152	9.9	288	2	US-08-525-940-15		Sequence 15, Appli
15	152	9.9	288	2	US-08-976-838-15		Sequence 15, Appli
16	142.5	9.3	568	1	US-07-862-0218-14		Sequence 14, Appli
17	142.5	9.3	568	5	PCF-US93-03164-14		Sequence 14, Appli
18	142.5	9.3	807	1	US-07-862-0218-10		Sequence 10, Appli
19	142.5	9.3	807	1	US-08-313-2888-10		Sequence 10, Appli
20	142.5	9.3	807	4	US-09-132-769-5		Sequence 5, Appli
21	142.5	9.3	807	5	PCF-US93-03164-10		Sequence 10, Appli
22	139.5	9.1	366	3	US-08-857-076-103		Sequence 103, App
23	139.5	9.1	486	3	US-08-746-559A-5		Sequence 5, Appli
24	139.5	9.1	516	3	US-08-746-559A-4		Sequence 4, Appli
25	139.5	9.1	807	4	US-09-132-769-1		Sequence 1, Appli
26	139.5	9.1	807	4	US-09-640-173-186		Sequence 186, App
27	139.5	9.1	807	4	US-09-713-550-186		Sequence 186, App

QY 76 CLSSCPGYYGTRYPDINKCTCKVDCTCFNK-NFCTKCKSGFYHLGKLDSCPEGL 134  
Db 780 CVTLCPAGFYADE--SQRNCKLCHPSCKKCVDBPEKCTVCKGFSLAGSGCIPDCBPGTY 837  
QY 135 ANNHTEC-----VSIVHCEAS-----EWS-----PWSP 158  
Db 838 FDSELRGCECHTGTGCVGGRBECIHCAKNFHFHDMKVCVPACGEGFYPEEMPGLPHKV 897  
QY 159 CMKKGKTC-----CFKRG-----TETVRDILQHPKSAKGNLCPPTSETRTCTVORK 206  
Db 898 CRRCDENCLSCAGSSRNCCKTGFTQGTGTCITNHTCSNADETFCEMVKSNRLC--ERK 955

## RESULT 2

US-08-447-642-2  
; Sequence 2, Application US/08447642  
; Patent No. 5989890  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; APPLICANT: KIEFER, MICHAEL C  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND  
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/447,642  
; FILING DATE: 23-MAY-1995  
; CLASSIFICATION: 424  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/284,941  
; FILING DATE: 2 August 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEELEY PH.D., RICHARD L.  
; REGISTRATION NUMBER: 30092  
; REFERENCE/DOCKET NUMBER: CHIR-009/01US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 843-5070  
; TELEFAX: (415) 857-0663  
; TELEX: 380816 COOLEY PA  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 969 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-447-642-2

Query Match 11.7%; Score 179; DB 2; Length 969;  
Best Local Similarity 23.3%; Pred. No. 1.2e-07;  
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;  
QY 18 YIGSQNASRGRQRMRHNPVNSQCGGCATCSD--YNGCLSKRPLFPVLERIGMKQIGV 75  
Db 737 YFGDTAARRCER-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779  
QY 76 CLSSCPGYYGTRYPDINKCTCKVDCTCFNK-NFCTKCKSGFYHLGKLDSCPEGL 134  
Db 780 CVTLCPAGFYADE--SQRNCKLCHPSCKKCVDBPEKCTVCKGFSLAGSGCIPDCBPGTY 837  
QY 135 ANNHTEC-----VSIVHCEAS-----EWS-----PWSP 158  
Db 838 FDSELRGCECHTGTGCVGGRBECIHCAKNFHFHDMKVCVPACGEGFYPEEMPGLPHKV 897

QY 159 CMKKGKTC-----CFKRG-----TETVRDILQHPKSAKGNLCPPTSETRTCTVORK 206  
Db 898 CRRCDENCLSCAGSSRNCCKTGFTQGTGTCITNHTCSNADETFCEMVKSNRLC--ERK 955

## RESULT 3

US-09-236-503-2  
; Sequence 2, Application US/09236503  
; Patent No. 6277590  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; APPLICANT: KIEFER, MICHAEL C  
; TITLE OF INVENTION: Compositions and Methods for PACE 4 and 4.1 Gene and  
; TITLE OF INVENTION: Polypeptides in Cells  
; FILE REFERENCE: CHIR-009/04US  
; CURRENT APPLICATION NUMBER: US/09/236,503  
; EARLIER FILING DATE: 1999-01-25  
; EARLIER APPLICATION NUMBER: 08/447,642  
; EARLIER FILING DATE: 1995-05-23  
; EARLIER APPLICATION NUMBER: 08/284,941  
; EARLIER FILING DATE: 1994-08-02  
; EARLIER APPLICATION NUMBER: 07/848,629  
; EARLIER FILING DATE: 1992-03-09  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 969  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-236-503-2

Query Match 11.7%; Score 179; DB 3; Length 969;  
Best Local Similarity 23.3%; Pred. No. 1.2e-07;  
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;  
QY 18 YIGSQNASRGRQRMRHNPVNSQCGGCATCSD--YNGCLSKRPLFPVLERIGMKQIGV 75  
Db 737 YFGDTAARRCER-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779  
QY 76 CLSSCPGYYGTRYPDINKCTCKVDCTCFNK-NFCTKCKSGFYHLGKLDSCPEGL 134  
Db 780 CVTLCPAGFYADE--SQRNCKLCHPSCKKCVDBPEKCTVCKGFSLAGSGCIPDCBPGTY 837  
QY 135 ANNHTEC-----VSIVHCEAS-----EWS-----PWSP 158  
Db 838 FDSELRGCECHTGTGCVGGRBECIHCAKNFHFHDMKVCVPACGEGFYPEEMPGLPHKV 897  
QY 159 CMKKGKTC-----CFKRG-----TETVRDILQHPKSAKGNLCPPTSETRTCTVORK 206  
Db 898 CRRCDENCLSCAGSSRNCCKTGFTQGTGTCITNHTCSNADETFCEMVKSNRLC--ERK 955

## RESULT 4

PCT-US93-02147A-2  
; Sequence 2, Application PC/TUS9302147A  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; APPLICANT: KIEFER, MICHAEL C  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND  
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/02147A  
FILING DATE: 19930309  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,629  
FILING DATE: 09-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: NESLEY PH.D., RICHARD L  
REGISTRATION NUMBER: 30092  
REFERENCE/DOCKET NUMBER: CHIR-009/0005  
TELEPHONE: (415) 494-7622  
TELEFAX: (415) 857-0663  
TELEX: 380816 COOLEY PA  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 969 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-02147A-2

Query Match 11.7% Score 179; DB 5; Length 969;  
Best Local Similarity 23.3%; Pred. No. 1.2e-07;  
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;

QY 18 YIGSONASRRQRMRHPNVSOQGGCATCSD--YNGCLSKCPRLFFVLIRIGMKQIGV 75  
DB 737 YFGDTAARCR-----CHKGCETSSRAATQCLSCR-RGFY-----HHQEMNT 779  
QY 76 CLSCSPSGYTYGTRYPDKTKVKVDCDTPNK-NFTCKKSGPYLHLGKCLDSCPRGLE 134  
DB 780 CVTLCPAGPYADE--SQKNCLKCHPSCKKCVDEPKCTVCKEGFSLARGSCIPDCERGTY 837  
QY 135 ANNHTMBC-----VSIVHCRAS-----EWS-----PWSP 158  
DB 838 FDELIRGCEHTTCTGTCVGPGRBECIHCAKFNHFDWKVCPACGEGFYPEEMFGLPHKV 897  
QY 159 CMKKGKTC-----GFGRG-----TETVRDILQHPSAKGNLCPPTSETRTCIQVRK 206  
DB 898 CRRCDENCLSCAGSRRNCRCKTGFTQLGTGTCITNHTCSNADETFCBMVKSRLC--ERK 955

RESULT 5  
US-08-368-852-15  
Sequence 15, Application US/08368852  
Patent No. 5691183  
GENERAL INFORMATION:  
APPLICANT: Franzusoff, Alex  
APPLICANT: Miranda, Luis R.  
TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES  
TITLE OF INVENTION: ENCODING SAID PROTEASES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 15  
FILING DATE: 05-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: 2848-11  
TELEPHONE: 303/863-9700  
TELEFAX: 303/863-0223  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-368-852-15

Query Match 10.1% Score 155; DB 1; Length 288;  
Best Local Similarity 28.2%; Pred. No. 4e-06;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATCSDYNGCLSKCPRLFFVLIRIGMKQIGVCLSCSPSGYTYGTRYPDI 92  
DB 13 PECSEVGCDDGPDPCND---CL---HYYYKLA---NNTRICVSSCPPGHY---HADK 58  
QY 93 NKCTKCKVDCDTPNK--NFTCKKSGPYLH--LGKCLDSCPRGLEANNHTMBCVSIHVC 148  
DB 59 KRCRKCAPNCSECFSGSHGDCMCKYGYFLNEETNSCVTHCPDGSYQDTKKMLC----- 112  
QY 149 EASWSPWSPCMKKGKTC-GFKCTETVRDILQHPSAKGNLCPPTSE 197  
DB 113 -----RKSENKKTCTBFHKCTCRDGLSLQ-----GSRCSVSCB 147

RESULT 6  
US-08-525-940-23  
Sequence 23, Application US/08525940  
Patent No. 5866351  
GENERAL INFORMATION:  
APPLICANT: Franzusoff, Alex  
APPLICANT: Miranda, Luis R.  
APPLICANT: Wolf, Joseph R.  
TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES  
TITLE OF INVENTION: ENCODING SAID PROTEASES  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525,940  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/368,852  
FILING DATE: 01-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/088,322  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2848-11-C1  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 799 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-525-940-23

Query Match 10.0%; Score 154; DB 2; Length 799;

Best Local Similarity 28.2%; Pred. No. 1.6e-05;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

Qy 35 PNVSQ-GCOQ-GCATCSYNGCLSCPKRLPFVLERIGMKQIGVCLSSCPGSGYGYTRYDPI 92  
Db 524 PECSEVCGDGPDPHCND---CL-----HYTYKLR---NNTRICVSSCPPGHY---HADK 569  
Qy 93 NKCTKCKVDCTCFNK--NFCTKCKSGFYHL--LGKCLDSCPGLNHNTHMTCVSIHVC 148  
Db 570 KRCRKCAPNCESCGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 623  
Qy 149 EASEWSPWPMCKKGKTC-GFKRGCTETVRDILQHPGSAKGNLCPPTSE 197  
Db 624 -----RKCSNCKTCTBFHNTCTCRDGLSLQ-----GSRCSVSCB 658

## RESULT 7

US-08-976-838-23

Sequence 23, Application US/08976838

Patent No. 5981259

GENERAL INFORMATION:

APPLICANT: Franzusoff, Alex

TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID

TITLE OF INVENTION: MOLECULES

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln St., Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/976,838

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2848-11-C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 799 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-976-838-23

Query Match 10.0%; Score 154; DB 2; Length 799;

Best Local Similarity 28.2%; Pred. No. 1.6e-05;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

Qy 35 PNVSQ-GCOQ-GCATCSYNGCLSCPKRLPFVLERIGMKQIGVCLSSCPGSGYGYTRYDPI 92  
Db 524 PECSEVCGDGPDPHCND---CL-----HYTYKLR---NNTRICVSSCPPGHY---HADK 569  
Qy 93 NKCTKCKVDCTCFNK--NFCTKCKSGFYHL--LGKCLDSCPGLNHNTHMTCVSIHVC 148  
Db 570 KRCRKCAPNCESCGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 623

Qy 149 EASEWSPWPMCKKGKTC-GFKRGCTETVRDILQHPGSAKGNLCPPTSE 197  
Db 624 -----RKCSNCKTCTBFHNTCTCRDGLSLQ-----GSRCSVSCB 658

## RESULT 8

US-08-525-940-21

Sequence 21, Application US/08525940

Patent No. 5866351

GENERAL INFORMATION:

APPLICANT: Franzusoff, Alex

APPLICANT: Miranda, Luis R.

APPLICANT: Wolf, Joseph R.

TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES

TITLE OF INVENTION: ENCODING SAID PROTEASES

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/525,940

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/368,852

FILING DATE: 01-JAN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/088,322

FILING DATE: 07-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2848-11-C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 881 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-525-940-21

Query Match 10.0%; Score 154; DB 2; Length 881;

Best Local Similarity 28.2%; Pred. No. 1.8e-05;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

Qy 35 PNVSQ-GCOQ-GCATCSYNGCLSCPKRLPFVLERIGMKQIGVCLSSCPGSGYGYTRYDPI 92  
Db 606 PECSEVCGDGPDPHCND---CL-----HYTYKLR---NNTRICVSSCPPGHY---HADK 651  
Qy 93 NKCTKCKVDCTCFNK--NFCTKCKSGFYHL--LGKCLDSCPGLNHNTHMTCVSIHVC 148  
Db 652 KRCRKCAPNCESCGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 705  
Qy 149 EASEWSPWPMCKKGKTC-GFKRGCTETVRDILQHPGSAKGNLCPPTSE 197  
Db 706 -----RKCSNCKTCTBFHNTCTCRDGLSLQ-----GSRCSVSCB 740

## RESULT 9

US-08-976-838-21



; Sequence 21, Application US/08976838  
; Patent No. 5981259  
; GENERAL INFORMATION:  
; APPLICANT: Franzusoff, Alex  
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln St., Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,838  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2848-11-C2  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 881 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-976-838-21

Query Match 10.0%; Score 154; DB 2; Length 881;  
Best Local Similarity 28.2%; Pred. No. 1.8e-05;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCQG-GCATCSYNGCLSCPKLPFLVLERIGHKQIGVCLSSCPGSGYGYTRYDPI 92  
Db 606 PECSEVGCDGPGDHCND---CL---HYVYKLGK---NNTRICVSSCPPGHY---HADK 651

QY 93 NKCTCKKVDCTCFNK--NFTCKCKSGFYHL--LGKCLDSCPEGLANNHTMBCVSIHVC 148  
Db 652 KRCRCKAPNCSCFGSHGDCMCKYGFLEETNSCVTHCPDGSYQDTKKNLC----- 705

QY 149 EASEWSPWSPCKKGTCTGFKRGTTETRVDRDILQHPSAKGNLCPPTSE 197  
Db 706 -----RKSENCKTCTFHNCETCRDGLSLQ-----GSRCSVSCE 740

RESULT 10  
US-08-525-940-18  
; Sequence 18, Application US/08525940  
; Patent No. 5866351  
; GENERAL INFORMATION:  
; APPLICANT: Franzusoff, Alex  
; APPLICANT: Miranda, Luis R.  
; APPLICANT: Wolf, Joseph R.  
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES  
; TITLE OF INVENTION: ENCODING SAID PROTEASES  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/525,940  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/368,852  
; FILING DATE: 01-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/088,322  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2848-11-C1  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-525-940-18

Query Match 10.0%; Score 154; DB 2; Length 915;  
Best Local Similarity 28.2%; Pred. No. 1.9e-05;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCQG-GCATCSYNGCLSCPKLPFLVLERIGHKQIGVCLSSCPGSGYGYTRYDPI 92  
Db 640 PECSEVGCDGPGDHCND---CL---HYVYKLGK---NNTRICVSSCPPGHY---HADK 685

QY 93 NKCTCKKVDCTCFNK--NFTCKCKSGFYHL--LGKCLDSCPEGLANNHTMBCVSIHVC 148  
Db 686 KRCRCKAPNCSCFGSHGDCMCKYGFLEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKKGTCTGFKRGTTETRVDRDILQHPSAKGNLCPPTSE 197  
Db 740 -----RKSENCKTCTFHNCETCRDGLSLQ-----GSRCSVSCE 774

RESULT 11  
US-08-976-838-18  
; Sequence 18, Application US/08976838  
; Patent No. 5981259  
; GENERAL INFORMATION:  
; APPLICANT: Franzusoff, Alex  
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln St., Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,838  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

```
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-18

Query Match      10.0%; Score 154; DB 2; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLGK---NNTRICVSSCPPGHY---HADK 685

QY 93 NKCTCKVDCTCFNK--NFTCKKSGFYH--LGKCLDSCPEGLEANNHTECVSIVHC 148
DB 686 KRCRKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNTCTECRDGLSLQ-----GSRCSVSCE 774

RESULT 12
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTRAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match      10.0%; Score 154; DB 4; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLGK---NNTRICVSSCPPGHY---HADK 685

QY 93 NKCTCKVDCTCFNK--NFTCKKSGFYH--LGKCLDSCPEGLEANNHTECVSIVHC 148
DB 686 KRCRKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNTCTECRDGLSLQ-----GSRCSVSCE 774

RESULT 12
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTRAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match      10.0%; Score 154; DB 4; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLGK---NNTRICVSSCPPGHY---HADK 685

QY 93 NKCTCKVDCTCFNK--NFTCKKSGFYH--LGKCLDSCPEGLEANNHTECVSIVHC 148
DB 686 KRCRKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNTCTECRDGLSLQ-----GSRCSVSCE 774

RESULT 13
US-09-214-555B-7
; Sequence 7, Application US/09214555B
```

```
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTRAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-7

Query Match      10.0%; Score 154; DB 4; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLGK---NNTRICVSSCPPGHY---HADK 685

QY 93 NKCTCKVDCTCFNK--NFTCKKSGFYH--LGKCLDSCPEGLEANNHTECVSIVHC 148
DB 686 KRCRKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNTCTECRDGLSLQ-----GSRCSVSCE 774

RESULT 14
US-08-525-940-15
; Sequence 15, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-525-940-15

Query Match 9.9%; Score 152; DB 2; Length 288;  
Best Local Similarity 27.7%; Pred. No. 7.4e-06;  
Matches 48; Conservative 24; Mismatches 53; Indels 48; Gaps 12;  
QY 35 PNVSQ-GCQG-GCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPGYYGTRYPDI 92  
DB 13 PECSEVGDGPGPHDND---CL---HYVYKLGK---NNTRICVSSCPPGHY---HADK 58  
QY 93 NKCTKCKVCDTCTCNK--NFCTKCKSGPYLH--LGKCLDSCPEGLEANNHTMECVSI VHC 148  
DB 59 KRCRCAPNCESCFCGSHGDCQMSCKYGYFLNBETNSCVTHCPDGSYQDTKKNLC----- 112  
QY 149 EASEWSPWSPCKMKGKTCGFRGTE-----TVRDILQHPSAKGNLCPPTSE 197  
DB 113 -----RKCSNFKTC-----TEFHICTCRDGLSL-----QGSRCVSCE 147

RESULT 15  
US-08-976-838-15  
Sequence 15, Application US/08976838  
Patent No. 5981259  
GENERAL INFORMATION:  
APPLICANT: Franzusoff, Alex  
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/976,838  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2848-11-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-976-838-15

Query Match 9.9%; Score 152; DB 2; Length 288;  
Best Local Similarity 27.7%; Pred. No. 7.4e-06;  
Matches 48; Conservative 24; Mismatches 53; Indels 48; Gaps 12;  
QY 35 PNVSQ-GCQG-GCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPGYYGTRYPDI 92

DB 13 PECSEVGDGPGPHDND---CL---HYVYKLGK---NNTRICVSSCPPGHY---HADK 58  
QY 93 NKCTKCKVCDTCTCNK--NFCTKCKSGPYLH--LGKCLDSCPEGLEANNHTMECVSI VHC 148  
DB 59 KRCRCAPNCESCFCGSHGDCQMSCKYGYFLNBETNSCVTHCPDGSYQDTKKNLC----- 112  
QY 149 EASEWSPWSPCKMKGKTCGFRGTE-----TVRDILQHPSAKGNLCPPTSE 197  
DB 113 -----RKCSNFKTC-----TEFHICTCRDGLSL-----QGSRCVSCE 147

Search completed: June 29, 2004, 17:04:01  
Job time : 14.5595 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 17:02:26 ; Search time 37.5837 Seconds

(without alignments)

2098.641 Million cell updates/sec

Title: US-09-894-912A-32

Perfect score: 1535

Sequence: 1 MHLRLISCFPIILFMFYIG.....QKGRARDKQKSVSVTVH 279

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1535	100.0	279	10	US-09-894-912A-32
2	1315.5	85.7	272	10	Sequence 32, Appl
3	1315.5	85.7	272	10	Sequence 10, Appl
4	1315.5	85.7	272	10	Sequence 34, Appl
5	1315.5	85.4	273	10	Sequence 23, Appl
6	1293	84.2	292	9	Sequence 13, Appl
7	1293	84.2	292	10	Sequence 166, App
8	1293	84.2	292	14	Sequence 48, Appl
9	1285	83.7	265	10	Sequence 25, Appl
10	1222.5	79.6	251	10	Sequence 26, Appl
11	1187	77.3	225	14	Sequence 16, Appl
12	1135	73.9	239	12	Sequence 4, Appl
13	1100	71.7	239	15	Sequence 1782, Ap
14	1045	68.1	195	15	Sequence 172, App
15	1041	67.8	195	15	Sequence 176, App
					Sequence 174, App

16	1018	66.3	180	12	US-10-087-192-1779	Sequence 1779, Ap
17	759	49.4	160	10	US-09-894-912A-14	Sequence 14, Appl
18	759	49.4	160	12	US-10-276-774-1744	Sequence 1744, Ap
19	648	42.2	263	14	US-10-125-852-18	Sequence 18, Appl
20	640.5	41.7	243	14	US-10-125-852-21	Sequence 21, Appl
21	632	41.0	265	14	US-10-125-852-24	Sequence 24, Appl
22	629	41.0	229	10	US-09-894-912A-25	Sequence 25, Appl
23	616.5	40.2	236	15	US-10-108-260A-4829	Sequence 4829, Ap
24	571.5	37.2	243	14	US-10-185-770-2	Sequence 2, Appl
25	569.5	37.1	243	14	US-10-125-852-13	Sequence 13, Appl
26	569.5	37.1	243	16	US-10-467-042-12	Sequence 12, Appl
27	565	36.8	250	14	US-10-125-852-3	Sequence 3, Appl
28	536	34.9	222	14	US-10-125-852-15	Sequence 15, Appl
29	531.5	34.6	229	14	US-10-125-852-6	Sequence 6, Appl
30	403.5	26.3	190	12	US-10-087-192-873	Sequence 873, App
31	370.5	24.1	161	12	US-10-087-192-876	Sequence 876, App
32	267	17.4	46	10	US-09-894-912A-18	Sequence 18, Appl
33	241.5	15.7	131	14	US-10-125-852-9	Sequence 9, Appl
34	208	13.6	110	14	US-10-125-852-11	Sequence 11, Appl
35	193	12.6	37	10	US-09-894-912A-20	Sequence 20, Appl
36	187	12.2	1548	14	US-10-180-903-2	Sequence 2, Appl
37	181.5	11.8	43	14	US-10-125-852-7	Sequence 7, Appl
38	179	11.7	969	10	US-09-961-403-6	Sequence 6, Appl
39	170.5	11.1	942	15	US-10-369-493-5070	Sequence 5070, Ap
40	150	9.8	42	10	US-09-894-912A-22	Sequence 22, Appl
41	145.5	9.5	827	15	US-10-369-493-5073	Sequence 5073, Ap
42	144.5	9.4	2471	15	US-10-190-115-27	Sequence 27, Appl
43	144.5	9.4	2471	15	US-10-369-072-27	Sequence 27, Appl
44	142.5	9.3	807	14	US-10-205-194-133	Sequence 133, App
45	140.5	9.2	660	12	US-09-939-853A-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1

US-09-894-912A-32  
; Sequence 32, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894, 912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266, 614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215, 733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757, 562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543, 774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 32  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-894-912A-32

Query Match 100.0%; Score 1535; DB 10; Length 279;  
Best Local Similarity 100.0%; Pred. No. 1.2e-114;  
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cyt 1 MHLRLISCFPIILFMFYIGSQNASRGRQRMPNVSQGGCATCDYNGCLSCRP 60  
Db 1 MHLRLISCFPIILFMFYIGSQNASRGRQRMPNVSQGGCATCDYNGCLSCRP 60  
Qy 61 LFFVLEIRIGHKQIGVCLLSCFSPGGYGYTRYPDINKCTKCKVDCDTCFKNKFNCTKCKSGPYL 120

Db 61 LPFVLERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120  
QY 121 HLKCLDSCPEGLEANNHTECVSIHVCHSESPSPCMKKGKTCGFKGTETVRDIL 180  
Db 121 HLKCLDSCPEGLEANNHTECVSIHVCHSESPSPCMKKGKTCGFKGTETVRDIL 180  
QY 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGRGRKRRKRLKLNKBERKETSSSDSK 240  
Db 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGRGRKRRKRLKLNKBERKETSSSDSK 240  
QY 241 GLESSIETPDQENKRRQKRRARDKQKSVSVSTVH 279  
Db 241 GLESSIETPDQENKRRQKRRARDKQKSVSVSTVH 279

## RESULT 2

US-09-894-912A-10  
; Sequence 10, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-894-912A-10

Query Match 85.7%; Score 1315.5; DB 10; Length 272;  
Best Local Similarity 87.1%; Pred. No. 3.8e-97;  
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60  
Db 1 MHLRLISWLPILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60  
QY 61 LPFVLERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120  
Db 61 LPFVLERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120  
QY 121 HLKCLDSCPEGLEANNHTECVSIHVCHSESPSPCMKKGKTCGFKGTETVRDIL 180  
Db 121 HLKCLDSCPEGLEANNHTECVSIHVCHSESPSPCMKKGKTCGFKGTETVRDIL 180  
QY 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGRGRKRRKRLKLNKBERKETSSSDSK 240  
Db 181 QHPSA--KGNLCPPTNETRKTIVQRKCKSGRGRGRKRRKRLKLNKBERKETSSSDSK 240  
QY 241 GLESSIETPDQENKRRQKRRARDKQKSVSVSTVH 279  
Db 237 SLESSKEIPEQRENK--QQKKRKVDK-QKSVSVSTVH 272

## RESULT 3

US-09-894-912A-34  
; Sequence 34, Application US/09894912A

; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 34  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-894-912A-34

Query Match 85.7%; Score 1315.5; DB 10; Length 272;  
Best Local Similarity 87.1%; Pred. No. 3.8e-97;  
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60  
Db 1 MHLRLISWLPILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60  
QY 61 LPFVLERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120  
Db 61 LPFVLERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120  
QY 121 HLKCLDSCPEGLEANNHTECVSIHVCHSESPSPCMKKGKTCGFKGTETVRDIL 180  
Db 121 HLKCLDSCPEGLEANNHTECVSIHVCHSESPSPCMKKGKTCGFKGTETVRDIL 180  
QY 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGRGRKRRKRLKLNKBERKETSSSDSK 240  
Db 181 QHPSA--KGNLCPPTNETRKTIVQRKCKSGRGRGRKRRKRLKLNKBERKETSSSDSK 240  
QY 241 GLESSIETPDQENKRRQKRRARDKQKSVSVSTVH 279  
Db 237 SLESSKEIPEQRENK--QQKKRKVDK-QKSVSVSTVH 272

## RESULT 4

US-10-125-852-23  
; Sequence 23, Application US/10125852  
; Publication No. US20030032034A1

US-10-125-852-23

; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1  
; FILE REFERENCE: HYS-43A  
; CURRENT APPLICATION NUMBER: US/10/125,852  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: US 60/316,368  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 23  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-125-852-23

Query Match 85.7%; Score 1315.5; DB 14; Length 272;  
Best Local Similarity 87.1%; Pred. No. 9.8e-97;  
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Qy 1 MHLRLISCFPIILNFMFYGISQNASRRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMFYGISQNASRRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60

Qy 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYL 120  
Db 61 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCTCFNKNFCTCKSGFYL 120

Qy 121 HLKCLDSCPEGLNANNHMECVSIHVCEASEWSPSCMKGGKTCGPKRGTTETVRDIL 180  
Db 121 HLKCLDNCPEGLNANNHMECVSIHVCESEWNPSPCTKGGTKCGPKRGTTETVRREII 180

Qy 181 QHPSAKGKGNLCPTSETRTCTIVQRKCKSGRGGKGRKRKRLNKKERKETSSESSDSK 240  
Db 181 QHPSA--KGNLCPTNETKCTVQRKCKQGBRGKGRKRKRPKNKESKE--AIPDSK 236

Qy 241 GLESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279  
Db 237 SLESSKEIPEQRENK--QOQKRRKVQDK--QKSVSVSTVH 272

## RESULT 5

US-09-894-912A-13  
; Sequence 13, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIORITY FILING DATE: 2002-05-10  
; PRIORITY FILING DATE: 2001-04-05  
; PRIORITY FILING DATE: 2001-04-05  
; PRIORITY FILING DATE: 2001-02-05  
; PRIORITY FILING DATE: 2001-02-05  
; PRIORITY FILING DATE: 2000-06-28  
; PRIORITY FILING DATE: 2000-06-28  
; PRIORITY FILING DATE: 2001-01-09  
; PRIORITY FILING DATE: 2001-01-09  
; PRIORITY FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-13

Query Match 85.4%; Score 1310.5; DB 10; Length 273;  
Best Local Similarity 87.1%; Pred. No. 9.6e-97;  
Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Qy 2 HRLRLISCFPIILNFMFYGISQNASRRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 61  
Db 3 HRLRLISWLFILNFMFYGISQNASRRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 62

Qy 62 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYL 121  
Db 63 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCTCFNKNFCTCKSGFYL 122

Qy 122 LGKCLDSCPEGLNANNHMECVSIHVCEASEWSPSCMKGGKTCGPKRGTTETVRDILQ 181  
Db 123 LGKCLDNCPEGLNANNHMECVSIHVCESEWNPSPCTKGGTKCGPKRGTTETVRREIIQ 182

Qy 182 QHPSAKGKGNLCPTSETRTCTIVQRKCKSGRGGKGRKRKRLNKKERKETSSESSDSK 241  
Db 181 QHPSA--KGNLCPTNETKCTVQRKCKQGBRGKGRKRKRLNKKERKETSSESSDSK 240

Db 183 HPSA--KGNLCPTNETKCTVQRKCKQGBRGKGRKRKRLNKKERKETSSESSDSK 238

Qy 242 LESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279  
Db 239 LESSKEIPEQRENK--QOQKRRKVQDK--QKSVSVSTVH 273

RESULT 6  
US-09-745-763-166  
; Sequence 166, Application US/09745763  
; Patent No. US20020065394A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; McCoy, John M.  
; Lavallie, Edward R.  
; Collins-Racie, Lisa A.  
; Evans, Cheryl  
; Merberg, David  
; Treacy, Maurice  
; Spaulding, Vikki  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; NUMBER OF SEQUENCES: 219  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/745,763  
; FILING DATE: 18-Jun-2000  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 166:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 292 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 166:  
US-09-745-763-166

Query Match 84.2%; Score 1293; DB 9; Length 292;  
Best Local Similarity 85.1%; Pred. No. 2.6e-95;  
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;

Qy 1 MHLRLISCFPIILNFMFYGISQNASRRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMFYGISQNASRRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60

Qy 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYL 120  
Db 61 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCTCFNKNFCTCKSGFYL 120

Qy 121 HLKCLDSCPEGLNANNHMECVSIHVCEASEWSPSCMKGGKTCGPKRGTTETVRDIL 180  
Db 121 HLKCLDNCPEGLNANNHMECVSIHVCESEWNPSPCTKGGTKCGPKRGTTETVRREII 180

Qy 181 QHPSAKGKGNLCPTSETRTCTIVQRKCKSGRGGKGRKRKRLNKKERKETSSESSDSK 240  
Db 181 QHPSA--KGNLCPTNETKCTVQRKCKQGBRGKGRKRKRLNKKERKETSSESSDSK 240

Db 181 QHPSA--KGNLCPTNETRKTQVKKCKQGERGKGRKRKPKNKGSKB--AIPDSK 236  
Qy 241 GLESSIETPDQENKRRARDKQKSVSVS 276  
Db 237 SLESSKEIPEQRENK--QOQKRRKQVQDKSGIEVT 270

## RESULT 7

US-09-894-912A-48  
; Sequence 48, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 48  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-48

Query Match 84.2%; Score 1293; DB 10; Length 292;  
Best Local Similarity 85.1%; Pred. No. 2.6e-95;  
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;  
Qy 1 MHLRLISCFPIILNFMEYIGSNASRRGRORRRHPNVSVSQCGGCATCSYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMEYIGSNASRRGRORRRHPNVSVSQCGGCATCSYNGCLSKPR 60  
Qy 61 LFFVLERIGMKQIGVCLSSCPSGYGTRYDPDKCKVDCDTCFNNKPFCTCKSGFY 120  
Db 61 LFFALERIGMKQIGVCLSSCPSGYGTRYDPDKCKVDCDTCFNNKPFCTCKSGFY 120  
Qy 121 HLGKLDSCPEGLEANNHTMECVSIHVCHASSEWSPWSPCKMKGKTCGFKGTETVRDIL 180  
Db 121 HLGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTCKGTCGFKGTETVRREII 180  
Qy 181 QHPSAKGNLCPTSTTRTCIVORKKCSGERGKGRKRKPKNKGSKB--AIPDSK 240  
Db 181 QHPSA--KGNLCPTNETRKTQVKKCKQGERGKGRKRKPKNKGSKB--AIPDSK 236  
Qy 241 GLESSIETPDQENKRRARDKQKSVSVS 276  
Db 237 SLESSKEIPEQRENK--QOQKRRKQVQDKSGIEVT 270

## RESULT 8

US-10-125-852-25  
; Sequence 25, Application US/10125852  
; Publication No. US20030032034A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-L  
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES  
; FILE REFERENCE: HYS-43A  
; CURRENT APPLICATION NUMBER: US/10/125,852  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; BEST FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-125-852-25

Query Match 84.2%; Score 1293; DB 14; Length 292;  
Best Local Similarity 85.1%; Pred. No. 2.6e-95;  
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;  
Qy 1 MHLRLISCFPIILNFMEYIGSNASRRGRORRRHPNVSVSQCGGCATCSYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMEYIGSNASRRGRORRRHPNVSVSQCGGCATCSYNGCLSKPR 60  
Qy 61 LFFVLERIGMKQIGVCLSSCPSGYGTRYDPDKCKVDCDTCFNNKPFCTCKSGFY 120  
Db 61 LFFALERIGMKQIGVCLSSCPSGYGTRYDPDKCKVDCDTCFNNKPFCTCKSGFY 120  
Qy 121 HLGKLDSCPEGLEANNHTMECVSIHVCHASSEWSPWSPCKMKGKTCGFKGTETVRDIL 180  
Db 121 HLGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTCKGTCGFKGTETVRREII 180  
Qy 181 QHPSAKGNLCPTSTTRTCIVORKKCSGERGKGRKRKPKNKGSKB--AIPDSK 240  
Db 181 QHPSA--KGNLCPTNETRKTQVKKCKQGERGKGRKRKPKNKGSKB--AIPDSK 236  
Qy 241 GLESSIETPDQENKRRARDKQKSVSVS 276  
Db 237 SLESSKEIPEQRENK--QOQKRRKQVQDKSGIEVT 270

## RESULT 9

US-09-894-912A-26  
; Sequence 26, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-26

Query Match 83.7%; Score 1285; DB 10; Length 265;  
Best Local Similarity 86.7%; Pred. No. 1e-94;  
Matches 234; Conservative 12; Mismatches 18; Indels 6; Gaps 3;  
Qy 1 MHLRLISCFPIILNFMEYIGSNASRRGRORRRHPNVSVSQCGGCATCSYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMEYIGSNASRRGRORRRHPNVSVSQCGGCATCSYNGCLSKPR 60



Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGTYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYL 120  
Db 61 LFFALERIGMKQIGVCLSSCPGSGYGTYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYL 120  
Qy 121 HLGKCLDSCPEGLNANNHTECVSIHVCEASWSPCMKGGKTCGPKRGTTETVRDIL 180  
Db 121 HLGKCLDNCPEGLNANNHTECVSIHVCEVSEWSPMPCCKGKTCGPKRGTTETVRBII 180  
Qy 181 QHPSAKGNLCPPPTSETTCTVQRKCKSGRGGKRRKRLKLNKBERKETSSSSDSK 240  
Db 181 QHPSA--KGNLCPPPTNETRKTCTVQRKCKGGRGGKRRKRLKPNKGSKE--AIPDSK 236  
Qy 241 GLESSIETPDQENKERQKQKRRARDKQ 270  
Db 237 SLESSKEIPEQRENK--QKQKRVQDKOK 264

RESULT 10  
US-09-894-912A-16  
; Sequence 16, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 16  
; LENGTH: 251  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-16

Query Match 79.6%; Score 1222.5; DB 10; Length 251;  
Best Local Similarity 86.8%; Pred. No. 9.3e-90;  
Matches 224; Conservative 11; Mismatches 16; Indels 7; Gaps 4;  
Qy 22 QNAGRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKRLFPVLRIGMKQIGVCLSSCP 81  
Db 1 QNAGRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKRLFPVLRIGMKQIGVCLSSCP 60  
Qy 82 SGYGTYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYLHLGKCLDSCPEGLNANNHTE 141  
Db 61 SGYGTYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYLHLGKCLDNCPEGLNANNHTE 120  
Qy 142 CVSIHVCEASWSPCMKGGKTCGPKRGTTETVRDILQHPSAKGNLCPPPTSETTCTC 201  
Db 121 CVSIHVCEVSEWSPMPCCKGKTCGPKRGTTETVRBIIQHPSA--KGNLCPPPTNETRKC 178  
Qy 202 IVQRKCKSGRGGKRRKRLKLNKBERKETSSSSDSKGLSSSIETPDQENKERQKQ 261  
Db 179 TVQRKCKQGRGGKRRKRLKPNKGSKE--AIPDSKLSLESSKEIPEQRENK--QKQK 234  
Qy 262 KRRARDKQKSVSVSTVH 279  
Db 235 KRVQDK--QKSVSVSTVH 251

RESULT 11  
US-10-185-770-4

; Sequence 4, Application US/10185770  
; Publication No. US20030022217A1  
; GENERAL INFORMATION:  
; APPLICANT: CECCARDI, Toni et al.  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; FILE REFERENCE: CL0001247  
; CURRENT APPLICATION NUMBER: US/10/185,770  
; CURRENT FILING DATE: 2002-07-01  
; PRIOR APPLICATION NUMBER: 60/301,852  
; PRIOR FILING DATE: 2001-07-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-185-770-4  
Query Match 77.3%; Score 1187; DB 14; Length 225;  
Best Local Similarity 92.5%; Pred. No. 5.7e-87;  
Matches 210; Conservative 5; Mismatches 10; Indels 2; Gaps 1;  
Qy 1 MHLRLISCFPIILNFMVEYIGSONASRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 60  
Db 1 MHLRLISWLFPIILNFMVEYIGSONASRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 60  
Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGTYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYL 120  
Db 61 LFFALERIGMKQIGVCLSSCPGSGYGTYPDINKCTKCKVDCDTCFNKNFCTKCKSGPYL 120  
Qy 121 HLGKCLDSCPEGLNANNHTECVSIHVCEASWSPCMKGGKTCGPKRGTTETVRDIL 180  
Db 121 HLGKCLDNCPEGLNANNHTECVSIHVCEVSEWSPMPCCKGKTCGPKRGTTETVRBII 180  
Qy 181 QHPSAKGNLCPPPTSETTCTVQRKCKSGRGGKRRKRLKLNK 227  
Db 181 QHPSA--KGNLCPPPTNETRKTCTVQRKCKGGRGGKRRKRLKPNK 225

RESULT 12  
US-10-087-192-1782  
; Sequence 1782, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1782  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-087-192-1782

Query Match 73.9%; Score 1135; DB 12; Length 239;  
Best Local Similarity 90.1%; Pred. No. 8.6e-83;  
Matches 201; Conservative 6; Mismatches 14; Indels 2; Gaps 1;  
Qy 1 MHLRLISCFPIILNFMVEYIGSONASRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 60  
Db 18 MHLRLISWLFPIILNFMVEYIGSONASRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 77



```

; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 176
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-094-886-176

Query Match      68.1%; Score 1045; DB 15; Length 195;
Best Local Similarity 92.8%; Pred. No. 1.1e-75;
Matches 181; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

QY 20 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPFLERIGMKQIGVCLSS 79
Db 1 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPFLERIGMKQIGVCLSS 60

QY 80 CPSGYGTYPDINKCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLANNHT 139
Db 61 CPSGYGTYPDINKCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLANNHT 120

QY 140 MECVSIHVCEASEWSPWCMKKGKTCGFKRGTTETVRDILQHPSAKGNLCPTTSETR 199
Db 121 MECVSIHVCEASEWSPWCMKKGKTCGFKRGTTETVRDILQHPSAKGNLCPTTSETR 178

QY 200 TCIVQRKKCKSGERG 214
Db 179 KCTVQRKKCKSGERG 193

RESULT 15
US-10-094-886-174
; Sequence 174, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Esha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennnda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie

```

```

; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderna, Steven
; APPLICANT: Larocheille, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD.
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 174
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-094-886-174

```

```

Query Match      67.8%; Score 1041; DB 15; Length 195;
Best Local Similarity 92.3%; Pred. No. 2.2e-75;
Matches 180; Conservative 6; Mismatches 7; Indels 2; Gaps 1;

QY 20 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPFLERIGMKQIGVCLSS 79
Db 1 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPFLERIGMKQIGVCLSS 60

QY 80 CPSGYGTYPDINKCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLANNHT 139
Db 61 CPSGYGTYPDINKCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLANNHT 120

QY 140 MECVSIHVCEASEWSPWCMKKGKTCGFKRGTTETVRDILQHPSAKGNLCPTTSETR 199
Db 121 MECVSIHVCEASEWSPWCMKKGKTCGFKRGTTETVRDILQHPSAKGNLCPTTSETR 178

QY 200 TCIVQRKKCKSGERG 214
Db 179 KCTVQRKKCKSGERG 193

```

Search completed: June 29, 2004, 17:15:57  
Job time : 37.5837 secs





Db 1162 -----PCQEGCKTC 1170  
||| : |||

RESULT 7  
I52527  
PACE4A - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999  
C:Accession: I52527  
R:Hosaka, M.; Murakami, K.; Nakayama, K.  
Biomed. Res. 15, 383-390, 1994  
A:Title: PACE4A is a ubiquitous endoprotease that has similar but not identical substrate specificities to the other members of the trypsin family.  
A:Reference number: I52527  
A:Accession: I52527  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-932 <RES>  
A:Cross-references: GB:I50060; NID:g769700; PIDN:BA08777.1; PID:g769701  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
F:172-410/Domain: subtilisin homology <SBT>

Query Match 11.1%; Score 171; DB 2; Length 932;  
Best Local Similarity 22.9%; Pred. No. 0.00073;  
Matches 55; Conservative 25; Mismatches 88; Indels 72; Gaps 12;

QY 18 YIGSQNASGRQRMRHPNVSGQGGGATCSQDYN--GCLCKPRLFFVLRIQGMKQIGV 75  
DB 700 YFGDAARRCR-----CHKGCETCTGRSPAQLSCR-RGFY-----HHQETNT 742  
QY 76 CLASCPSGYGYTRYPDINKTKKVDCTCFNK-NFCTKCKGFFYLHLGKLDSCPEGLE 134  
DB 743 CVTLCPAGLYADEBSQRL--CLURHPSCQKVDPEKCTVCKEGFSLARGSCIPDCEPGTY 800  
QY 135 ANNHTEEC-----VSIVHCEAS-----EWS-----PWSP 158  
DB 801 FDSSELYKQGECHTTCRTCVGPSREECIHCAKSFHFQDMKCVPCGSGFYPEEMGLPHKV 860  
QY 159 CMKKGKTC-----GPFRG-----TETVRDILQHPSAKGKGNLCPTSETTCTIVORK 206  
DB 861 CRRCEENCLSCGSSRNCRCKAGTQIGTSCITNHTYCSNADETFCFEMVKSNNRLC--ERK 918

RESULT 8  
D87803  
protein bli-4D [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 21-Jun-2002  
C:Accession: D87803  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: D87803  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-942 <STO>  
A:Cross-references: GB:chr\_I; PIDN:AAB96754.1; PID:g2773243; GSPDB:GN00019  
C:Genetics:  
A:Gene: bli-4D  
A:Map position: 1  
C:Superfamily: kexin; subtilisin homology

Query Match 11.1%; Score 170.5; DB 2; Length 942;  
Best Local Similarity 28.3%; Pred. No. 0.0008;  
Matches 53; Conservative 18; Mismatches 75; Indels 41; Gaps 9;

QY 36 NVSGQGGGATCSQDYNGLCSCKPRLFFVLRIQGMKQIGVCLSSCPSGYGYTRYPDINKC 95  
DB 674 NCHDEKNGCTESSATSSCPACK-HLTOTLRNKGSGP-KCVQKCDDTY-----YLDGDKC 727

QY 96 TKKVDCTCFNKNFCTCKSGPYL-----HL--GKCLDSCPEGLRANHTMBCUSIV 146  
Db 728 RMCSSCHCTTCAEVCTCPGSLLLIDVDNMPHYDHGCKVSCPPGLVADYBS-----NLV 783  
QY 147 HCEASEWSPWSPCKGKTCGPKRGTEVRDILQHPKSAKGNLCPPYSETRTC--IVQ 204  
Db 784 QAK-----CIWRKDLGGDG-----YYINAVGKCDLDCSSCTCTAGPMS 823  
QY 205 RKCKSKG 211  
Db 824 CERCKSKG 830  
RESULT 9  
B48225  
probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat  
N/Alternate names: PC5 precursor  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 23-Feb-1997  
C/Accession: B48225  
R.;Jussion, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993  
A/Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate pro-  
A/Reference number: A48225; MUID:93342056; PMID:8341687  
A/Accession: B48225  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-915 <LUS>  
C/Cross-references: GB:114933  
C/Suprafamily: subtilisin-like proteinase PACE4; subtilisin homology  
F:1-34/Domain: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase  
F:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>  
F:117-915/Product: probable proprotein convertase 5 #status experimental <MAT>  
F:164-402/Domain: subtilisin homology <SST>  
F:173,214,388/Active site: Asp, His, Ser #status predicted  
Query Match 10.7%; Score 164; DB 2; Length 915;  
Best Local Similarity 27.1%; Pred. No. 0.002;  
Matches 48; Conservative 20; Mismatches 57; Indels 48; Gaps 9;  
QY 35 PNVSQ--GCOQ-----GCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPGYYGTRY 89  
Db 640 PCESEVGCDGPGPDHCTCLHYHKLNNTR-----LCVSSCPGHP---H 682  
QY 90 PDINKCTCKKVDCTCFNKNP--CTKCKSGPYLH--LGRKCLDSCPEGLRANHTMBCVSI 145  
Db 683 ADKRCRCKCAPNCESCFGSHADQCLSKYGYFLNBTSCVAQCPEGSYQDIKKNIC--- 739  
QY 146 VHCEASEWSPWSPCKKGGKTC-GPKRGTEVRDILQHPKSAKGNLCPPYTSR 197  
Db 740 -----GKCSNCKTCTGFHNCTCKGGLSLQ-----GSRCSVTCE 774  
RESULT 10  
A48225  
subtilisin-like proprotein convertase (EC 3.4.21.-) PC5 precursor - mouse  
N/Alternate names: kexin homolog; serine proteinase PC6  
C/Species: Mus musculus (house mouse)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C/Accession: A48225; JX0248  
R.;Jussion, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993  
A/Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate pro-  
A/Reference number: A48225; MUID:93342056; PMID:8341687  
A/Accession: A48225  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-915 <LUS>  
C/Cross-references: GB:114932; NID:9293327; PIDN:AAA74636.1; PID:9293328  
J.Nakagawa, T.; Hosaka, M.; Torii, S.; Watanabe, T.; Murakami, K.; Nakayama, K.  
J. Biochem. 113, 132-135, 1993  
A/Title: Identification and functional expression of a new member of the mammalian Kex2-

A/Reference number: JX0248; MUID:93224489; PMID:8468318  
A/Accession: JX0248  
A/Molecule type: mRNA  
A/Residues: 1-915 <NAK>  
C/Cross-references: DDBJ:D12619; NID:9220565; PIDN:BA02143.1; PID:9220566  
A/Note: the authors translated the codon GGC for residue 915 as Ala  
C/Suprafamily: subtilisin-like proteinase PACE4; subtilisin homology  
C/Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-116/Domain: propeptide #status predicted <PRO>  
F:117-915/Product: proprotein convertase PC5 #status experimental <MAT>  
F:164-402/Domain: subtilisin homology <SST>  
F:173,214,388/Active site: Asp, His, Ser #status predicted  
Query Match 10.7%; Score 163.5; DB 1; Length 915;  
Best Local Similarity 29.1%; Pred. No. 0.0022;  
Matches 44; Conservative 18; Mismatches 50; Indels 39; Gaps 8;  
QY 20 GSNAISRGRQRMRHNPVNSQCGCGCATCSDYNG--CLSCPKRLFFVLERIGMKQIGVCL 77  
Db 679 GHYHADK-KRCRCAPN-----CESCFGSGDQCLSKCYGYFL-----NEETSSCV 723  
QY 78 SSCPSGYGYTRYPDINK--CTKCKVDCTCFNKNFCTCKSGPYLHGLKCLDSCPEGLEA 135  
Db 724 TQCPDG---SYEDIKKNVCGKCSNCKACIGFHNCTCKGGLSLQSGSRCSVTCDGQFF 779  
QY 136 NNHTMECVSIVHCEASEWSPWSPCKKGGKTC 166  
Db 780 NGEH--DC-----QPCRFPCATC 794  
RESULT 11  
T37314  
probable kexin (EC 3.4.21.61) - Caenorhabditis elegans (fragment)  
N/Alternate names: bliesterase 4  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C/Accession: T37314  
R.;Thacker, C.; Peters, K.; Srayko, M.; Rose, A.M.  
Genes Dev. 9, 956-971, 1995  
A/Title: The bli-4 locus of Caenorhabditis elegans encodes structurally distinct kex2/5  
A/Reference number: Z21679; MUID:95293228; PMID:7774813  
A/Accession: T37314  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-570 <THA>  
C/Cross-references: EMBL:L29440; NID:9459702; PIDN:AAA98752.1; PID:9459703  
C/Genetics:  
A/Gene: bli-4  
A/Map position: 1  
C/Keywords: alternative splicing; hydrolase; serine proteinase  
Query Match 10.6%; Score 162; DB 2; Length 570;  
Best Local Similarity 27.1%; Pred. No. 0.0019;  
Matches 54; Conservative 21; Mismatches 76; Indels 48; Gaps 10;  
QY 31 RMHNPVNSQ-----COGCGCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPG 83  
Db 290 RGSFPDLTSGWKLSCDCEGCGCTESSATSCFAFK-HLTOTLRNKGSGP-KCVQKDDT 347  
QY 84 YGTRYPDINKCTKCKVDCTCFNKNFCTCKSGPYL-----HL--GKCLDSCPEGLE 134  
Db 348 Y-----YLDGDKCKMCKSSHCHCTCTKAETCTCPGSLLLIDVDNMPHYDHGCKVSCPPGLV 403  
QY 135 ANNHTMECVSIVHCEASEWSPWSPCKKGGKTCGPKRGTEVRDILQHPKSAKGNLCPP 194  
Db 404 ADYES---NLVQAK-----CIWRKDLGGDG-----YYINAVGKCDLDCS 439  
QY 195 TSETRTC--IVQRCKCKSKG 211  
Db 440 SCSTCTAPGPMSCCKCKSKG 458



RESULT 12  
Gene PACE4 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 22-Jun-1999  
C:Accession: I53282  
R:Johnson, R.C.; Darlington, D.N.; Hand, T.A.; Bloomquist, B.T.; Mains, R.E.  
Endocrinology 135, 1178-1185, 1994  
A:Title: PACE4: a subtilisin-like endoprotease prevalent in the anterior pituitary and  
A:Reference number: I53282; MUID:94349873; PMID:8070361  
A:Accession: I53282  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-937 <RES>  
A:Cross-references: GB:L31894; NID:g496221; PIDN:AAA61987.1; PID:g496222  
C:Genetics:  
A:Gene: PACE4  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
P:177-415/Domain: subtilisin homology <SET>  
Query Match 10.5%; Score 161; DB 2; Length 937;  
Best Local Similarity 22.5%; Pred. No. 0.0032;  
Matches 54; Conservative 24; Mismatches 90; Indels 72; Gaps 12;  
QY 18 YISQNASRRQRHNPVNSQCGGCGATCSQDYN--GCLSCKPLFPVLERIGMKQIGV 75  
DB 705 YFGDTAARRCR-----CHKGCCTGRTGSPQCLSCR-RGFY-----HHQSTNT 747  
QY 76 CLSSCPGYYGTRYPDINKCTCKKVDCTCFNK-NFCTCKSGFYHLGKCLDSCPEGLE 134  
DB 748 CVTLCPAGLYADESQL--CLRCHPSQKQVDBPEKSTVCKEGESLARGSCIPDCBPGTY 805  
QY 135 ANNHTMEC-----VSVIHCAS-----EWS-----PMSP 158  
DB 806 FDSRLRGCHGCHTCTCVGPSREECIHCAKSFHQDWKVPACGGFYEPPEMPLPHKV 865  
QY 159 CMKKGKTC-----GFKRG-----TETVRDILQHPSAKGNKGLCPPTSETRTICVORK 206  
DB 866 CRRCDENLCSGSSRNCRKAGTQLTGTCITNHTCSNADETFCFEMVKSRLC--BRK 923  
RESULT 13  
G02428  
subtilisin-like proprotein convertase (EC 3.4.21.-) 5 precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 17-Mar-2000  
C:Accession: G02428  
R:Reudelhuber, T.L.  
Submitted to the EMBL Data Library, February 1996  
A:Reference number: H01242  
A:Accession: G02428  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-899 <REU>  
A:Cross-references: EMBL:U49114; NID:gl218057; PIDN:AAA91807.1; PID:gl218058  
C:Genetics:  
A:Gene: PCS  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: hydrolase; serine proteinase  
P:148-386/Domain: subtilisin homology <SET>  
Query Match 10.0%; Score 154; DB 2; Length 899;  
Best Local Similarity 28.2%; Pred. No. 0.0089;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;  
QY 35 PNVSQ-GCOG-GCATCSQDYNGCLSCKPLFPVLERIGMKQIGVCLSSCPGYYGTRYDPI 92  
DB 624 PECSEVCGDGPDPHCND---CL-----HYTYKLR-----NNTRICVSSCPGPHY---HADK 669  
QY 93 NKCTCKKVDCTCFNK--NFCTCKSGFYHL--LGKCLDSCPEGLANNHTMECVSIHVC 148  
DB 670 KRCKKAPNCESCFGSHGDCMSCKGYFLNEETNSCVTHCPDGSYQDTKNLC----- 723

QY 149 EASEWSPWSPCMKKGKTC-GFKRGTYETVRDILQHPSAKGNKGLCPPTSE 197  
DB 724 -----RKSENCKTCTEFHNCTECRDGLSLQ-----GSRCSVSCR 758  
RESULT 14  
JC6148  
subtilisin-like proprotein convertase (EC 3.4.21.-) homolog - human  
N:Alternate names: PC6A protease  
C:Species: Homo sapiens (man)  
C:Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 22-Jun-1999  
C:Accession: JC6148  
R:Miranda, L.; Wolf, J.; Pichuanter, S.; Duke, R.; Franzusoff, A.  
Proc. Natl. Acad. Sci. U.S.A. 93, 7695-7700, 1996  
A:Title: Isolation of the human PC6 gene encoding the putative host protease for HIV-1  
A:Reference number: JC6148; MUID:96353880; PMID:8755538  
A:Contents: CEM T-cell  
A:Accession: JC6148  
A:Molecule type: mRNA  
A:Residues: 1-915 <MIR>  
A:Cross-references: GB:U56387; NID:gl498312; PIDN:RAC50643.1; PID:gl498313  
C:Comment: This protein functions as a soluble enzyme within the secretory pathway. It  
C:Genetics:  
A:Gene: pc6A  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase  
P:164-402/Domain: subtilisin homology <SET>  
Query Match 10.0%; Score 154; DB 2; Length 915;  
Best Local Similarity 28.2%; Pred. No. 0.009;  
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;  
QY 35 PNVSQ-GCOG-GCATCSQDYNGCLSCKPLFPVLERIGMKQIGVCLSSCPGYYGTRYDPI 92  
DB 640 PECSEVCGDGPDPHCND---CL-----HYTYKLR-----NNTRICVSSCPGPHY---HADK 685  
QY 93 NKCTCKKVDCTCFNK--NFCTCKSGFYHL--LGKCLDSCPEGLANNHTMECVSIHVC 148  
DB 686 KRCKKAPNCESCFGSHGDCMSCKGYFLNEETNSCVTHCPDGSYQDTKNLC----- 739  
QY 149 EASEWSPWSPCMKKGKTC-GFKRGTYETVRDILQHPSAKGNKGLCPPTSE 197  
DB 740 -----RKSENCKTCTEFHNCTECRDGLSLQ-----GSRCSVSCR 774  
RESULT 15  
A33837  
insulin-like growth factor I receptor precursor - rat  
N:Contains: insulin-like growth factor I alpha chain; insulin-like growth factor I beta  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 16-Mar-1990 #sequence\_revision 03-Nov-1995 #text\_change 21-Jul-2000  
C:Accession: JC2461; A33837; PC1131  
R:Pedrini, M.T.; Giorino, F.; Smith, R.J.  
Biochem. Biophys. Res. Commun. 202, 1038-1046, 1994  
A:Title: cDNA cloning of the rat IGF I receptor: Structural analysis of rat and human IGI  
A:Reference number: JC2461; MUID:94324926; PMID:8048916  
A:Accession: JC2461  
A:Molecule type: mRNA  
A:Residues: 1-1371 <PED>  
R:Warner, H.; Woloschak, M.; Adamo, M.; Shen-Orr, Z.; Roberts Jr., C.T.; Leroith, D.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7451-7455, 1989  
A:Title: Developmental regulation of the rat insulin-like growth factor I receptor gene.  
A:Reference number: A33837; MUID:90017496; PMID:2477843  
A:Accession: A33837  
A:Molecule type: mRNA  
A:Residues: 1-364 <WER>  
A:Cross-references: GB:M27293  
R:Kurachi, H.; Jobo, K.; Ohta, M.; Kawasaki, T.; Itoh, N.  
Biochem. Biophys. Res. Commun. 187, 934-939, 1992  
A:Title: A new member of the insulin receptor family, insulin receptor-related receptor,  
A:Reference number: PC1130; MUID:92412145; PMID:1530648  
A:Accession: PC1131



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:24 ; Search time 8.46481 Seconds  
(without alignments)  
1716.232 Million cell updates/sec

Title: US-09-894-912A-32

Perfect score: 1535

Sequence: 1 MHLRLISCFIILNFMIEYIG.....QKRRARDKQKSVSVSTVH 279

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187	12.2	1877	1	PCK5 MOUSE
2	181.5	11.8	1696	1	PCK5 BRACL
3	179	11.7	969	1	PAC4 HUMAN
4	174	11.3	1679	1	FUR2 DROME
5	164	10.7	1877	1	PCK5 RAT
6	162	10.6	943	1	BL14 CAREL
7	161	10.5	937	1	PAC4 RAT
8	154	10.0	913	1	PCK5 HUMAN
9	153.5	10.0	374	1	WIFI_XENLA
10	149	9.7	1370	1	IGIR RAT
11	145.5	9.5	803	1	FSPO XENLA
12	144.5	9.4	2471	1	NTC2 RAT
13	142.5	9.3	807	1	FSPO RAT
14	142	9.3	1373	1	IGIR MOUSE
15	140.5	9.2	2470	1	NTC2 MOUSE
16	140	9.1	378	1	WIFI BRARE
17	139.5	9.1	1367	1	IGIR HUMAN
18	139	9.1	2524	1	NOTC XENLA
19	138.5	9.0	379	1	WIFI HUMAN
20	136	8.9	379	1	WIFI MOUSE
21	136	8.9	687	1	VS41_GIALA
22	133.5	8.7	2482	1	VMP PIG
23	131	8.5	1297	1	IRR HUMAN
24	131	8.5	2146	1	INSR DROME
25	131	8.5	2471	1	NTC2 HUMAN
26	130.5	8.5	1382	1	INSR HUMAN
27	130	8.5	2531	1	NTC1 MOUSE
28	130	8.5	2531	1	NTC1 RAT
29	129.5	8.4	1372	1	INSR MOUSE
30	129.5	8.4	1383	1	INSR RAT
31	129	8.4	2437	1	NTC1 BRARE
32	129	8.4	2556	1	NTC1 HUMAN
33	128	8.3	461	1	TR1A_PIG

#### ALIGNMENTS

##### RESULT 1

PCK5 MOUSE  
ID PCK5\_MOUSE STANDARD; PRT; 1877 AA.  
AC Q04592; Q62040;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
DE (Protein convertase PCS) (Subtilisin/kexin-like protease PC5) (PC6)  
DE (Subtilisin-like proprotein convertase 6) (SPC6).  
GN PCSK5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PCSB).  
RC STRAIN=ICR; TISSUE=Intestine;  
RX MEDLINE=93327934; PubMed=8335106;  
RA Nakagawa T., Murakami K., Nakayama K.;  
RT "Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";  
RL FBS Lett. 327:165-171(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM PCSA).  
RC TISSUE=Brain, and Intestine;  
RX MEDLINE=93224489; PubMed=8468318;  
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;  
RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACE4.";  
RL J. Biochem. 113:132-135(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM PCSA).  
RC TISSUE=Adrenal cortex;  
RX MEDLINE=93342056; PubMed=8341687;  
RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
RN [4]  
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
RX MEDLINE=97103178; PubMed=8947550;  
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;  
RT "The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";  
RL J. Cell Biol. 135:1261-1275(1996).  
RN [5]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=96293359; PubMed=8698813;  
RA Constam D.B., Calfon M., Robertson E.J.;  
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

34 127 8.3 1300 1 IRR\_MOUSE  
35 124.5 8.1 667 1 TS11\_GIALA  
36 124 8.1 581 1 IRR\_RAT  
37 124 8.1 1300 1 IRR\_CAVPO  
38 123.5 8.0 375 1 DEK\_HUMAN  
39 123.5 8.0 1477 1 HTK7\_HYDAT  
40 123 8.0 2871 1 FBN1\_HUMAN  
41 120.5 7.9 2871 1 FBN1\_PIG  
42 119 7.8 1713 1 LFB1\_MOUSE  
43 118.5 7.7 833 1 SRC\_MOUSE  
44 118.5 7.7 1367 1 LT23\_CAEEL  
45 118 7.7 1170 1 TSP2\_BOVIN

Q9wt14 mus musculus  
Q03185 giardia lam  
Q64716 rattus norv  
P14617 cavia porce  
P35659 homo sapien  
Q25197 hydra atten  
P35555 homo sapien  
Q9t936 sus scrofa  
Q8c919 mus musculus  
P59222 mus musculus  
P24348 caenorhabdi  
Q95116 bos taurus

RL J. Cell Biol. 134:181-191(1996).  
 RN [6]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=97436919; PubMed=9291583;  
 RA Rancourt S.L., Rancourt D.E.;  
 RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic  
 RL implantation, somitogenesis, and skeletal formation";  
 Dev. Genet. 21:75-81(1997).  
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity  
 CC within the constitutive and regulated secretory pathway. Capable  
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible  
 CC for the maturation of gastrointestinal peptides. May be involved  
 CC in the cellular proliferation of adrenal cortex via the activation  
 CC of growth factors.  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -!- SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED  
 CC SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO  
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH  
 CC EARLY ENDOSOMES.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=PCSB; Synonyms=Long;  
 CC IsoId=Q04592-1; Sequence=Displayed;  
 CC Name=PCSA; Synonyms=Short;  
 CC IsoId=Q04592-2; Sequence=VSP\_005438;  
 CC -!- TISSUE SPECIFICITY: PCSA is expressed in most tissues but is most  
 CC abundant in the intestine and adrenals. PCSB is expressed in the  
 CC intestine, adrenals and lung but not in the brain.  
 CC -!- DEVELOPMENTAL STAGE: Weakly expressed throughout the embryo,  
 CC except in the developing nervous system, the ribs and the liver,  
 CC but markedly upregulated at discrete sites during development. At  
 CC E6.5, prominent expression observed in differentiated decidua. At  
 CC E7.5, intense expression in extraembryonic endoderm, amnion and  
 CC nascent mesoderm. At E8.5, abundant expression in somites and yolk  
 CC sac followed by a confinement to dermamyotome compartment. Between  
 CC E9.5 and E11.5, abundant expression in AER (thickened ectodermal  
 CC cells of limb buds). At E12.5, expression in the limbs is confined  
 CC to the condensing mesenchym surrounding the cartilage. At this  
 CC stage, strong expression also detected in vertebral and facial  
 CC cartilage primordia and in the muscle of the tongue. At E16.5  
 CC abundant expression in epithelial cells of the intestinal villi.  
 CC Isoform A is most abundant at all stages but significant levels of  
 CC isoform B occur at E12.5.  
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
 CC assisting the folding of the zymogen within the endoplasmic  
 CC reticulum.  
 CC -!- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain  
 CC sorting information. AC 1 directs TGN localization and interacts  
 CC with the TGN sorting protein PACS-1.  
 CC -!- SIMILARITY: Belongs to peptidase family 88.  
 CC -!- SIMILARITY: Contains 1 homo B/P domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D17583; BAA04507.1; -;  
 DR EMBL; D12619; BAA02143.1; -;  
 DR EMBL; L14932; AAA74636.1; -;  
 DR PIR; A48225; A48225.  
 DR PIR; S34583; S34583.  
 DR HSSP; Q99405; IMET.  
 DR MEROPS; S08.076; -;  
 DR MGD; MGI:97515; Pcbk5.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.

DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR002884; Peptidase\_S88.  
 DR InterPro; IPR009020; Protease\_inhib.  
 DR Pfam; PF01483; P\_protease\_1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PD00723; SUBTILISIN.  
 DR ProDom; PD000717; P\_domain; 1.  
 DR SMART; SM00261; FU; 22.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE SER; 1.  
 DR Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing;  
 KW Transmembrane.  
 FT SIGNAL 1 34  
 FT PROPEP 35 116  
 FT CHAIN 117 1877  
 FT TYPE 5.  
 FT PROTEIN CONVERTASE SUBTILISIN/KEXIN  
 FT DOMAIN 117 1768 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1769 1789 POTENTIAL.  
 FT DOMAIN 1790 1877 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 117 452 CATALYTIC.  
 FT DOMAIN 464 602 HOMO B.  
 FT DOMAIN 638 1753 CYS-RICH MOTIF (CRM) REGION.  
 FT DOMAIN 1825 1844 AC 1.  
 FT DOMAIN 1856 1877 AC 2.  
 FT SITE 116 117 CLEAVAGE (AUTO-) (BY SIMILARITY).  
 FT SITE 521 523 CELL ATTACHMENT SITE (POTENTIAL).  
 FT ACT\_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1220 1220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1317 1317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1711 1711 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1733 1733 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPPLIC 878 915  
 FT GYDIDQGHGCTCRASCAKCGWFTQDCISCPTVRVLD ->  
 FT ATBESWARGFCMLVKNNLCCQKVLQQLCKCTCFQG  
 FT (in isoform PCSA).  
 FT /FTid=VSP\_005438.  
 FT Missing (in isoform PCSA).  
 FT /FTid=VSP\_005439.  
 FT VARSPLIC 916 1877  
 FT SEQUENCE 1877 AA; 209287 MW; EC850E2DF20BA1C3 CRC64;  
 Query Match 12.2%; Score 187; DB 1; Length 1877;  
 Best Local Similarity 27.8%; Pred. No. 1.2e-05;  
 Matches 52; Conservative 21; Mismatches 72; Indels 42; Gaps 11;  
 QY 38 SQCCQGGCATCSYNG--CLSKPRLFFVLERIGMQLGVCLSSCSGGYGYTRPDI--N 93  
 DB 1209 NQPCSSCKTC---NGSLCASCPTGMVLMQ-----ACVPCPQG---TPWSTSG 1253  
 QY 94 KCTKCKVDCTCFNKNFCTKCKS---GFVLHGLKLDSCPEGLEANNHTMBCVSIHVCE 149  
 DB 1254 SCEKSEDCVSCGADLCCQCLSQPNTLLHGRGCHVSCPEGYAKGVCE-----HCS 1308  
 QY 150 ASEWSPWSPCKMKGKTCGPKRGRTETVRDIL-----QHPESAKGKMLCPPTSETRTIV 203  
 DB 1309 ----SPKTCGNATSCNSCBGDFVLDHGVCKWCTCEKHVAVEGCKHCP--ERCQCIH 1362  
 QY 204 QR--KKC 208  
 DB 1363 EKTCKEC 1369

RESULT 2  
ID PKCS\_BRACL STANDARD; PRT; 1696 AA.  
AC Q9NJ15; Q9NJ14; Q9NJ16;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE (Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
DE (Protein convertase PCG-like) (aPC6).  
GN PC6.  
OS Branchiostoma californiensis (California lancelet) (Amphioxus).  
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
OC Branchiostoma.  
OX NCBI\_TaxID=7738;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).  
RA MEDLINE=20175281; PubMed=10708868;  
RA Oliva A.A. Jr., Chan S.J., Steiner D.P.;  
RT "Evolution of the prohormone convertases: identification of a  
RL homologue of PC6 in the protochordate amphioxus.";  
RL Biochim. Biophys. Acta 1477:338-348(2000).  
CC -|- FUNCTION: Likely to represent a widespread endoprotease activity  
CC within the constitutive and regulated secretory pathway. Capable  
CC of cleavage at the RX(K/R)R consensus motif (By similarity).  
CC -|- CATALYTIC ACTIVITY: Release of mature proteins from their  
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg/-Zaa bonds, where Xaa  
CC can be any amino acid and Yaa is Arg or Lys.  
CC -|- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.  
CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.  
CC -|- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=B;  
CC IsoId=Q9NJ15-1; Sequence=Displayed;  
CC Name=A;  
CC IsoId=Q9NJ15-2; Sequence=VSP\_005444, VSP\_005445;  
CC Name=C;  
CC IsoId=Q9NJ15-3; Sequence=VSP\_005442, VSP\_005443;  
CC -|- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
CC assisting the folding of the zymogen within the endoplasmic  
CC reticulum.  
CC -|- SIMILARITY: Belongs to peptidase family S8.  
CC -|- SIMILARITY: Contains 1 homo B/P domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; AF184615; AAF26300.1; -;  
DR EMBL; AF184616; AAF26301.1; -;  
DR EMBL; AF184617; AAF26302.1; -;  
DR HSP; Q99405; IMPT.  
DR MEROPS; S08.UBP; -;  
DR InterPro; IPR006212; Furin repeat.  
DR InterPro; IPR009030; Grow fac recep.  
DR InterPro; IPR002029; Peptidase\_S8.  
DR InterPro; IPR002884; Peptidase\_S8.  
DR InterPro; IPR009020; Protease\_inhib.  
DR Pfam; PF01483; P\_protein; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR ProDom; PD000717; P\_domain; 1.  
DR SMART; SM00261; FU; 17.  
DR PROSITE; PS00136; SUBTILASE\_ASP; FALSE\_NEG.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
KW Cleavage on pair of basic residues; Repeat; Alternative splicing;  
KW Transmembrane.  
FT SIGNAL 1 25 POTENTIAL.

POTENTIAL.  
PROPROTEIN CONVERTASE SUBTILISIN/KEXIN  
TYPE 5.  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.  
CYTOPLASMIC (POTENTIAL).  
CATALYTIC.  
HOMO B.  
CYS-RICH MOTIF (CRM) REGION.  
CLEAVAGE (AUTO-) (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
DDTLDRGCTICGPGVMDRRKKKCKACHTQKCSDEY  
DDTCACTACNDGFLDADSCGACP -> AENQNDASFCPPA  
PREVSVLAELALGHLRYSLDTPQSNPPPTVLGADRRL  
TTATSAAGRA (in isoform C).  
/FTid=VSP\_005442.  
Missing (in isoform C).  
/FTid=VSP\_005443.  
CHPTCKESDEYDPTACNDGFLDADSCGACPPGQFL  
HGGDCDSCHREKTC -> IARCVDRRDRSCLDLVRNFC  
VRRYFVKRCCTCKLYMEDRRMRGSSQTOGRN (in  
isoform A).  
/FTid=VSP\_005444.  
Missing (in isoform A).  
/FTid=VSP\_005445.  
SEQUENCE 1696 AA; 188410 MW; 281CBEL784257CBD CRC64;  
Query Match 11.8%; Score 181.5; DB 1; Length 1696;  
Best Local Similarity 23.9%; Pred. No. 2.6e-05;  
Matches 68; Conservative 23; Mismatches 72; Indels 121; Gaps 15;  
Qy 33 MHPNVSQCGGCGATCS--DYNGCLSCP----- 59  
Db 1328 LHGGDCDSCHREKTCDDPHDNCSCQPSYLNDDQCSHCPGTPETVEDDSGETVL 1387  
Qy 60 --RLFFV-----LERIGMKQIGVCLSSCPSPGYGTRYPDI-NKCTCKV 100  
Db 1388 QCLCHVNCVTCGEGBEDCHCANDIKYQDGRCVTECQEGH---YPLUTNECQCS 1443  
Qy 101 DCDTPC--NRNPCTCKSGFYHLGKCLDSCPEGL-----EANN 137  
Db 1444 DCECTDGPNDQCVTCFYPNYLVLGKCLDCEPGEYDTMRQEGECGCHPSCATCNEGN 1503  
Qy 138 HTMECVSIHV-----CRASEW-----SPWSPCKKKGTCGFKRGTRVRDIL 180  
Db 1504 Y--NCLSCPYSGLGEGVCYPMCEHEYYVETQICECDNSCKTC---RG--STAHDCL 1556  
Qy 181 QHPSAKG---KGNLC-----PTSE-----TRTCIVQRK 206  
Db 1557 SCRAPYGHAMKHLCTACCEGSPENEYCCICHESTRLCLTDRE 1600  
RESULT 3  
PAC4\_HUMAN STANDARD; PRT; 969 AA.  
AC Q9UEJ7; Q9UEJ6; Q9UEJ5; Q9UEJ4; Q9UEJ3; Q9UEJ2; Q9UEJ1; Q9UEJ0;  
AC Q9UEJ9; Q9Y4G9; Q9Y4H0; Q9Y4H1;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)  
DE (Subtilisin/kexin-like protease FACE4) (Subtilisin-like proprotein  
DE convertase 4) (SPC4).  
GN FACE4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;

[1] SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).

TSUBS=Hepato, and kidney;

MEDLINE=92075167; PubMed=1741956;

Klefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D., Barr P.J.;

"Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15.";

DNA Cell Biol. 10:757-769(1991).

[2] SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).

TSUBS=Placenta;

MEDLINE=94235049; PubMed=8179631;

Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H., Matsuda Y.;

"Identification of novel cDNAs encoding human kexin-like protease, PACE4 isoforms.";

RT PACE4 isoforms.";

RT PACE4 isoforms.";

Biochem. Biophys. Res. Commun. 200:943-950(1994).

[3] ERATUM.

MEDLINE=95071480; PubMed=7980617;

Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H., Matsuda Y.;

"Identification of novel cDNAs encoding human kexin-like protease, PACE4 isoforms.";

RT PACE4 isoforms.";

Biochem. Biophys. Res. Commun. 204:1381-1382(1994).

[4] SEQUENCE FROM N.A. (ISOFORM PACE4A-II).

TSUBS=Placenta;

Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A., Matsuda Y.;

"Identification of a novel PACE4 isoform, PACE4B.";

Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

[5] SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).

TSUBS=Cerebellum;

MEDLINE=97335942; PubMed=9192737;

Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K., Akamatsu T., Nagamune H., Matsuda Y.;

"A novel human PACE4 isoform, PACE4E is an active processing protease containing a hydrophobic cluster at the carboxy terminus.";

J. Biochem. 121:941-948(1997).

[6] SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; PACE4A-II; PACE4C; PACE4D; PACE4E-I AND PACE4E-II).

MEDLINE=98021085; PubMed=9378725;

Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S., Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;

"Genomic organization and alternative splicing of human PACE4 (SPC4), kexin-like processing endoprotease.";

J. Biochem. 122:438-452(1997).

[7] ALTERNATIVE SPLICING (ISOFORM PACE4CS).

MEDLINE=97064242; PubMed=8906861;

Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;

"Functional analysis of human PACE4-A and PACE4-C isoforms: Identification of a new PACE4-CS isoform.";

FEBS Lett. 396:31-36(1996).

[8] CHARACTERIZATION.

MEDLINE=99233559; PubMed=10215603;

Sucic J.F., Moehring J.M., Innocencio N.M., Luchini J.W., Moehring T.J.;

"Endoprotease PACE4 is Ca2+-dependent and temperature-sensitive and can partly rescue the phenotype of a furin-deficient cell strain.";

Biochem. J. 339:639-647(1999).

[9] PROCESSING.

MEDLINE=98408849; PubMed=9738469;

Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K., Tsuji A., Matsuda Y.;

"Biothetic processing and quaternary interactions of proprotein convertase SPC4 (PACE4).";

FEBS Lett. 434:155-159(1998).

-1- FUNCTION: Likely to represent an endoprotease activity within the constitutive secretory pathway, with unique restricted distribution in both neuroendocrine and non-neuroendocrine tissues and capable of cleavage at the RX(K/R)R consensus motif.

-1- CATALYTIC ACTIVITY: Release of mature proteins from their proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys.

-1- COFACTOR: Calcium (Potential).

-1- SUBUNIT: The PACE4A-I precursor protein seems to exist in the reticulum endoplasmic as both a monomer and a dimer-sized complex whereas mature PACE4A-I exists only as a monomer, suggesting that propeptide cleavage affects its tertiary or quaternary structure.

-1- SUBCELLULAR LOCATION: PACE4A-I and PACE4A-II are secreted. PACE4C and PACE4CS are not secreted and remain probably in zymogen form in endoplasmic reticulum. PACE4E-I and PACE4E-II are retained intracellularly probably through a hydrophobic cluster in their C-terminus. PACE4B might be secreted.

-1- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=8;

Names=PACE4A-I; Synonyms=PACE4;

ISOIDS=P29122-1; Sequence=Displayed;

Names=PACE4A-II;

ISOIDS=P29122-2; Sequence=VSP\_005436;

Names=PACE4B; Synonyms=PACE4.1;

ISOIDS=P29122-3; Sequence=VSP\_005428, VSP\_005429;

Note=Probably enzymatically inactive;

Names=PACE4C;

ISOIDS=P29122-4; Sequence=VSP\_005432, VSP\_005433;

Note=Probably enzymatically inactive;

Names=PACE4CS;

ISOIDS=P29122-5; Sequence=VSP\_005430, VSP\_005431;

Note=Probably enzymatically inactive;

Names=PACE4D;

ISOIDS=P29122-6; Sequence=VSP\_005427, VSP\_005434, VSP\_005435;

Note=Probably enzymatically inactive;

Names=PACE4E-I;

ISOIDS=P29122-7; Sequence=VSP\_005437;

Names=PACE4E-II;

ISOIDS=P29122-8; Sequence=VSP\_005436, VSP\_005437;

TISSUE SPECIFICITY: Each PACE4 isoform exhibits a unique restricted distribution. PACE4A-I is expressed in heart, brain, placenta, lung, skeletal muscle, kidney, pancreas, but at comparatively higher levels in the liver. PACE4A-II is at least expressed in placenta. PACE4B was only found in the embryonic kidney cell line from which it was isolated. PACE4C and PACE4D are expressed in placenta. PACE4E-I is expressed in cerebellum, placenta and pituitary. PACE4E-II is at least present in cerebellum.

-1- DOMAIN: The propeptide domain acts as an intramolecular chaperone assisting the folding of the zymogen within the endoplasmic reticulum. Isoform PACE4D lacks the propeptide domain.

-1- SIMILARITY: Belongs to peptidase family S8.

-1- SIMILARITY: Contains 1 homo B/P domain.

-1- SIMILARITY: Contains 1 PLAC domain.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M80482; AAA5998.1; ..

EMBL; AB001914; BAA21620.1; ..

EMBL; AB001898; BAA21620.1; JOINED.

EMBL; AB001900; BAA21620.1; JOINED.

EMBL; AB001901; BAA21620.1; JOINED.

EMBL; AB001902; BAA21620.1; JOINED.

EMBL; AB001903; BAA21620.1; JOINED.

EMBL; AB001904; BAA21620.1; JOINED.

EMBL; AB001905; BAA21620.1; JOINED.

DR EMBL; AB001914; BAA21621.1; --  
 DR EMBL; AB001898; BAA21621.1; JOINED.  
 DR EMBL; AB001900; BAA21621.1; JOINED.  
 DR EMBL; AB001901; BAA21621.1; JOINED.  
 DR EMBL; AB001902; BAA21621.1; JOINED.  
 DR EMBL; AB001903; BAA21621.1; JOINED.  
 DR EMBL; AB001904; BAA21621.1; JOINED.  
 DR EMBL; AB001905; BAA21621.1; JOINED.  
 DR EMBL; AB001906; BAA21621.1; JOINED.  
 DR EMBL; AB001907; BAA21621.1; JOINED.  
 DR EMBL; AB001908; BAA21621.1; JOINED.  
 DR EMBL; AB001909; BAA21621.1; JOINED.  
 DR EMBL; AB001914; BAA21622.1; --  
 DR EMBL; AB001901; BAA21622.1; JOINED.  
 DR EMBL; AB001902; BAA21622.1; JOINED.  
 DR EMBL; AB001903; BAA21622.1; JOINED.  
 DR EMBL; AB001904; BAA21622.1; JOINED.  
 DR EMBL; AB001905; BAA21622.1; JOINED.  
 DR EMBL; AB001906; BAA21622.1; JOINED.  
 DR EMBL; AB001907; BAA21622.1; JOINED.  
 DR EMBL; AB001908; BAA21622.1; JOINED.  
 DR EMBL; AB001914; BAA21623.1; --  
 DR EMBL; AB001898; BAA21623.1; JOINED.  
 DR EMBL; AB001900; BAA21623.1; JOINED.  
 DR EMBL; AB001901; BAA21623.1; JOINED.  
 DR EMBL; AB001902; BAA21623.1; JOINED.  
 DR EMBL; AB001903; BAA21623.1; JOINED.  
 DR EMBL; AB001904; BAA21623.1; JOINED.  
 DR EMBL; AB001905; BAA21623.1; JOINED.  
 DR EMBL; AB001906; BAA21623.1; JOINED.  
 DR EMBL; AB001907; BAA21623.1; JOINED.  
 DR EMBL; AB001908; BAA21623.1; JOINED.  
 DR EMBL; AB001914; BAA21624.1; --  
 DR EMBL; AB001898; BAA21624.1; JOINED.  
 DR EMBL; AB001900; BAA21624.1; JOINED.  
 DR EMBL; AB001901; BAA21624.1; JOINED.  
 DR EMBL; AB001902; BAA21624.1; JOINED.  
 DR EMBL; AB001903; BAA21624.1; JOINED.  
 DR EMBL; AB001904; BAA21624.1; JOINED.  
 DR EMBL; AB001914; BAA21624.1; JOINED.  
 Query Match 11.7%; Score 179; DB 1; Length 969;  
 Best Local Similarity 23.3%; Pred. No. 2.3e-05;  
 Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;  
 Qy 18 YIGSNASRGRRQRMRHNVSGGGGATCSD--YNGCLSKPRLFPVLRIGMKQIGV 75  
 Db 737 YFGTAAARCR-----CHKGCTVSSRAATQCLSCR-RCFY-----HHQEWNT 779  
 Qy 76 CLSSCPGYYGTRYPDINKTKCKYDCTCFNK-NFTCKKSGFYHLHGKLDSCPEGLE 134  
 Db 780 CVTLCPAGFYADE--SQKNCLKCHPSCKKCVDEPEKCTVCKGFSLARGSCITPCDCEPTY 837  
 Qy 135 ANNHTEC-----VSIVCEAS-----EWS-----PWSF 158  
 Db 838 FDSLEIRGECHTGTGTCVGFREECIRCAKNFHFDMKVCVACGEGYPPEMPLPHKV 897  
 Qy 159 CMKKGKTC---GPKRG-----TETVRDILQHPSAKGKGNLCPTSETRTCTIVQRK 206  
 Db 398 CRRDCNCLSCAGSRNCSCKTGTGTQGTSCITNHTCSNADETCVMKSNRLC--ERK 955  
 RESULT 4  
 FUR2 DROME  
 AC P30432; Q24301; PRT; 1679 AA.  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DB Furin-like protease 2 precursor (BC 3.4.21.75) (Purin 2).  
 GN FUR2 OR CQ18734/CG4235.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RC STRAIN=Oregon-R, Tuebingen, and Iso-1;  
 CC RX MEDLINE=92381036; PubMed=1512259;  
 CC RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,  
 CC RT Rentrop M., Gateff E.A.F., Leunissen J.A.M., van de Ven W.J.M.;  
 CC "Cloning and functional expression of Dfurin2, a subtilisin-like  
 CC RT proprotein processing enzyme of Drosophila melanogaster with multiple  
 CC RT repeats of a cysteine motif";  
 CC RL J. Biol. Chem. 267:17208-17215(1992).  
 CC [2]  
 CC RN SEQUENCE FROM N.A.  
 CC RP STRAIN=ISO-1;  
 CC RX MEDLINE=95186060; PubMed=7880443;  
 CC RA Roebroek A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,  
 CC RT van de Ven W.J.M.;  
 CC "The Dfur2 gene of Drosophila melanogaster: genetic organization,  
 CC RT expression during embryogenesis, and pro-protein processing activity  
 CC RT of its translational product Dfurin2";  
 CC RL DNA Cell Biol. 14:233-234(1995).  
 CC [3]  
 CC RN SEQUENCE FROM N.A.  
 CC RP STRAIN=Berkely;  
 CC RX MEDLINE=20196006; PubMed=10731132;  
 CC RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 CC RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 CC RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 CC RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 CC RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 CC RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.B.G.,  
 CC RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 CC RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 CC RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 CC RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 CC RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 CC RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 CC RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 CC RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 CC RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 CC RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 CC RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 CC RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 CC RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 CC RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 CC RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 CC RA Laske P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 CC RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 CC RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 CC RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 CC RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 CC RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 CC RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 CC RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 CC RA Spier B., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
 CC RA Svirskas R., Tector C.C., Turner R., Venter E., Wang A.H., Wang X.,  
 CC RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 CC RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 CC RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 CC RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 CC RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 CC "The genome sequence of Drosophila melanogaster.";  
 CC RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: Purin is likely to represent the ubiquitous endoprotease  
 CC activity within constitutive secretory pathways and capable of  
 CC cleavage at the RX(R/R)R consensus motif (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,  
 CC complement component C3 and von Willebrand factor from their  
 CC respective precursors.



CC -1- TISSUE SPECIFICITY: Transient expression in a subset of central  
 CC nervous system neurons during embryonic stages 12-13. Expression  
 CC in developing tracheal tree from stage 13 to end of embryonic  
 CC development.  
 CC -1- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M94375; AAA28551.1; -;  
 CC EMBL; L33831; AAA69860.1; -;  
 CC EMBL; AE003502; AAF48598.1; -;  
 CC PIR; A43434; A43434.  
 CC HSSP; Q99405; IMPT.  
 CC MEROPS; S08.049; -;  
 CC FlyBase; FBgn0004598; Fur2.  
 CC GO; GO:0004276; F: furin activity; IDA.  
 CC InterPro; IPR006212; Furin repeat.  
 CC InterPro; IPR009030; Grow fac recep.  
 CC InterPro; IPR002029; Peptidase\_S8.  
 CC InterPro; IPR002884; Peptidase\_S8B.  
 CC InterPro; IPR009020; Protease\_inhib.  
 CC Pfam; PF01483; P\_protease; 2.  
 CC Pfam; PF00082; Peptidase\_S8; 1.  
 CC PRINTS; PR00723; SUBTILISIN.  
 CC ProDom; PD000717; P\_domain; 1.  
 CC SMART; SM00261; FU\_10.  
 CC PROSITE; PS00136; SUBTILASE ASP; 1.  
 CC PROSITE; PS00137; SUBTILASE HIS; 1.  
 CC PROSITE; PS00138; SUBTILASE SER; 1.  
 CC Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;  
 KW Multigene family; Zymogen; Repeat.  
 FT SIGNAL; 1 2  
 FT PROPEP 1 318 POTENTIAL.  
 FT CHAIN 319 1679 FURIN-LIKE PROTEASE 2.  
 FT ACT\_SITE 417 417 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 456 456 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DOMAIN 961 1443 10 X TANDEM REPEATS, CYS-RICH.  
 FT REPEAT 961 1006 1.  
 FT REPEAT 1007 1056 2.  
 FT REPEAT 1057 1103 3.  
 FT REPEAT 1104 1152 4.  
 FT REPEAT 1153 1204 5.  
 FT REPEAT 1205 1253 6.  
 FT REPEAT 1254 1298 7.  
 FT REPEAT 1299 1345 8.  
 FT REPEAT 1346 1392 9.  
 FT REPEAT 1393 1443 10.  
 FT TRANSMEM 1512 1532 POTENTIAL.  
 FT DOMAIN 1533 1679 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 927 927 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1060 1060 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1181 1181 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 152 153 MISSING (IN REF. 1).  
 FT CONFLICT 177 177 V -> F (IN REF. 1).  
 FT CONFLICT 213 213 V -> VDQL (IN REF. 1).  
 SQ SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CP6 CRC64;

Query Match 11.3%; Score 174; DB 1; Length 1679;  
 Best Local Similarity 24.5%; Pred. No. 8.6e-05;  
 Matches 48; Conservative 15; Mismatches 61; Indels 72; Gaps 6;  
 QY 23 NASRGRQRQRMHFN---VSQGGCGGCATCSDY--NGCLSKCKPRLPFVFLERIGMKQIGVCL 77  
 DB 994 NTCVSRCPSPSPFNQVGICWPCDHTCTCTCAGAGPDSCLTCAPAHLRVID-----LAVCL 1047  
 QY 78 SSCPSGY-----GTRYDPI 92  
 DB 1048 QFCPDGYFENRNRTVCPCPNCASQDHPEYCTSCDHLVMMHKKVCYACPLDTYETSD 1107  
 QY 93 NKCTCKKGVDCITCF--NKNFCTCKSGFYHLGLKCLDSCPEGLNANNHTWECVSIHVCEA 150  
 DB 1108 NKCAFCHSTCATNGPFDQCIYCRSRYSYAWQNKCLISCPDGFYADKKRLSCM----- 1160  
 QY 151 SEWSPWSPCMKGGKTC 166  
 DB 1161 -----PCQEGCKTC 1169  
 RESULT 5  
 PCK5 RAT  
 ID\_PCK5\_RAT STANDARD; PRT; 1877 AA.  
 AC P41I3; Q62914;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
 DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PCS)  
 DE (rPC5) (PC6) (Fragment).  
 GN PCSK5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.  
 RC TISSUE=Adrenal gland;  
 RX MEDLINE=93342056; PubMed=8341687;  
 RA Lussion J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
 RT "CDNA structure of the mouse and rat subtilisin/kexin-like PC5; a  
 RT candidate proprotein convertase expressed in endocrine and  
 RT nonendocrine cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
 RN [2]  
 RP SEQUENCE OF 1676-1877 FROM N.A. (ISOFORM LONG).  
 RC TISSUE=Adrenal gland;  
 RA De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=20214819; PubMed=10749928;  
 RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;  
 RT "The PC6B cytoplasmic domain contains two acidic clusters that direct  
 RT sorting to distinct trans-Golgi network/endosomal compartments.";  
 RL Mol. Biol. Cell 11:1257-1273(2000).  
 RN [4]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=97166043; PubMed=9013936;  
 RA Zheng M., Seidah N.G., Pintar J.E.;  
 RT "The developmental expression in the rat CNS and peripheral tissues of  
 RT proteases PC5 and PACE4 mRNAs: comparison with other proprotein  
 RT processing enzymes.";  
 RL Dev. Biol. 181:268-283(1997).  
 CC -1- FUNCTION: Likely to represent a widespread endoprotease activity  
 CC within the constitutive and regulated secretory pathway. Capable  
 CC of cleavage at the RX(K/R) consensus motif. May be responsible  
 CC for the maturation of gastrointestinal peptides. May be involved  
 CC in the cellular proliferation of adrenal cortex via the activation  
 CC of growth factors.  
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their





STRAIN=Sprague-Bawley; TISSUE=Hypothalamus, and Pituitary;  
MEDLINE=94339873; PubMed=8070361;  
RA Johnson R.C., Darling D.N., Hand T.A., Bloomquist B.T., Mains R.E.;  
RT "PACE": a subtilisin-like endoprotease prevalent in the anterior  
RL Endocrinology 135:1178-1185(1994).  
CC -I- FUNCTION: Likely to represent an endoprotease activity within the  
CC constitutive secretory pathway, with unique restricted  
CC distribution in both neuroendocrine and non-neuroendocrine tissues  
CC and capable of cleavage at the RK(K/R)R consensus motif.  
CC -I- CATALYTIC ACTIVITY: Release of mature proteins from their  
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
CC can be any amino acid and Yaa is Arg or Lys.  
CC -I- COPACTOR: Calcium (Potential).  
CC -I- TISSUE SPECIFICITY: High expression in the anterior pituitary and  
CC in several brain regions, the atrium, and the ventricle.  
CC -I- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
CC assisting the folding of the zymogen within the endoplasmic  
CC reticulum.  
CC -I- SIMILARITY: Belongs to peptidase family S8.  
CC -I- SIMILARITY: Contains 1 homo B/P domain.  
CC -I- SIMILARITY: Contains 1 PLAC domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: L31894; AAA61987.1; -.  
CC DR PIR: I53282; I53282.  
CC DR HSSP: Q93405; IMPT.  
CC DR MEROPS: S08.075; -.  
CC InterPro: IPR006212; Furin repeat.  
CC DR InterPro: IPR000209; Peptidase\_S8.  
CC DR InterPro: IPR002884; Peptidase\_S8B.  
CC DR InterPro: IPR003020; Protease\_Inhib.  
CC DR Pfam: PF01483; P\_protease\_1.  
CC DR Pfam: PF00082; Peptidase\_S8; 1.  
CC DR PRINTS: PR00723; SUBTILISIN.  
CC DR PRODOM: PD000717; P\_domain; 1.  
CC DR SMART: SM00261; FU; 5.  
CC DR PROSITE: PS50900; PLAC; 1.  
CC DR PROSITE: PS00135; SUBTILASE\_ASP; 1.  
CC DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
CC DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
CC DR K0 Hydroxylase: Serine protease; Glycoprotein; Zymogen; Signal; Calcium;  
CC KW Hydroxylase: Serine protease; Glycoprotein; Zymogen; Signal; Calcium;  
CC KW Cleavage on pair of basic residues; Repeat.  
CC FT SIGNAL 1 45  
CC FT PROPEP 45 132  
CC FT CHAIN 133 937  
-----  
CC FT DOMAIN 133 454  
CC FT DOMAIN 477 615  
CC FT DOMAIN 680 898  
CC FT DOMAIN 899 937  
CC FT SITE 132 133  
CC FT SITE 534 536  
CC FT ACT\_SITE 186 186  
CC FT ACT\_SITE 227 227  
CC FT ACT\_SITE 401 401  
CC FT CARBOHYD 240 240  
CC FT CARBOHYD 882 882  
CC FT CARBOHYD 900 900  
CC SEQUENCE 937 AA; 10405 MW; F3865557C33705C8 CRC64;  
-----  
Query Match 10.58; Score 161; DB 1; Length 937;  
Best local Similarity 22.58; Pred. No. 0.0004;  
Matches 54; Conservative 24; Mismatches 90; Indels 72; Gaps 12;  
QV 18 YIGSONASRRGRORRHHPNVSQCGGCATCSQDYN--GCLSCKPRLPFVFLRIGMKQIGV 75

Db 705 YFGDTAARRRR-----CHKGCETGSRPTQCLSCR-RGFY-----HHQSTNT 747  
Qy 76 CLSSCPGGYTRYPDINKTKCKVDDCDTCFNK-NFTCKSGGYLHLGKLDSCPEGLB 134  
Db 748 CVTLCPAGLYADBSQRL--CLRCHPSQCKCVDEPEKSTVCKEFGSLARGSCIPDCPGTY 805  
Qy 135 ANNHMTMC-----VSIHVCRAS-----EWS-----PSP 158  
Db 806 FDELIRGCEHHYTCRTCVGSPRECIHCASPFQDMKVPACGEGFYPEMPGLPHKV 865  
Qy 159 CMKKGKTC---GFRRG-----TETRVEDILQHPGSAKGNLCPTSTRTCTIVORK 206  
Db 866 CRRDCNCLSCGSSRNCRCRAGTQLTGTCITNHTCSNADFTFCMVKNRLC--SRK 923

RESULT 8

PKCS\_HUMAN

ID PKCS\_HUMAN STANDARD; PRT; 913 AA.

AC Q92824; Q13527;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Proteinkin convertase subilisin/kexin type 5 precursor (EC 3.4.21.-)  
DE (Proteinprotein convertase PCS) (Subilisin/kexin-like protease PCS) (PC6)  
DE (hPC6).  
GN PCSK5 OR PCS OR PC6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
[1]  
SEQUENCE FROM N.A.  
RC TISSUE=T-cell;  
RX MEDLINE=96353880; PubMed=8755538;  
RA Miranda L., Wolf J., Pichuanes S., Duke R., Franzusoff A.;  
RT "Isolation of the human PC6 gene encoding the putative host protease  
for HIV-1 gp160 processing in CD4+ T lymphocytes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700(1996).  
[2]  
REVIEWS.  
RA Franzusoff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 913 FROM N.A.  
RA Reudelhuber T.L.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Likely to represent a widespread endoprotease activity  
within the constitutive and regulated secretory pathway. Capable  
of cleavage at the RX(K/R) consensus motif.  
CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
proteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
can be any amino acid and Yaa is Arg or Lys.  
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
Event-Alternative splicing; Named isoforms=1;  
Comment=2 isoforms may be produced;  
Name=1;  
CC IsoId=Q92824-1; Sequence=Displayed;  
CC -!- TISSUE SPECIFICITY: Expressed in T-lymphocytes.  
CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
assisting the folding of the zymogen within the endoplasmic  
reticulum.  
CC -!- SIMILARITY: Belongs to peptidase family S8.  
CC -!- SIMILARITY: Contains 1 homo B/P domain.  
CC -!- SIMILARITY: Contains 1 PLAC domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announcement>)

Query Match  
Best Local  
Matches 54 18



FT DISULFID 289 291 POTENTIAL.  
 FT DISULFID 300 309 POTENTIAL.  
 FT DISULFID 305 315 POTENTIAL.  
 FT DISULFID 321 323 POTENTIAL.  
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 374 AA; 41071 MW; E26P973B0F00ACP8 CRC64;

Query Match 10.0%; Score 153.5; DB 1; Length 374;  
 Best Local Similarity 24.4%; Pred. No. 0.00056;  
 Matches 60; Conservative 23; Mismatches 78; Indels 85; Gaps 14;

QY 41 CQGCAT---CSDYNGC-----LCKPRLFFVLIRGMQIGVC-----LSSC 80  
 DB 177 CTGCGRNGFCNDKHVCECDPFGFYGPCEKALCMPRCM-----NGGLCVTPGLCIC 227  
 QY 81 PSYGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYHLGKLDSCPEGLEANNHTM 140  
 DB 228 PPGYGINCDKVNCTTHC-LNGGTCTF-----YPGKCI--CPSGYGE----- 266  
 QY 141 ECVSIVHCRASEWSFWSPCMKKGTCTGKGTETVRVDILQHPSAKG-KGNLC-----P 193  
 DB 267 -----QCTSKQ--QPCRNKGRKCKGNK-----CKSKGYQDLCCKPVCEP 307  
 QY 194 PTSTRTCIQVRK-KCSGKGRKGRKRRKGLMKERKETSSSSDKGLSSLETDPQQ 252  
 DB 308 SCGAHGTCIFBNKQCKCKEGWNG-----RYCNKKYGSMLMNLARTGSRNRQHTSPK 359  
 QY 253 ENKERQ 258  
 DB 360 RTEDRQ 365

## RESULT 10

IGIR RAT STANDARD; PRT; 1370 AA.  
 AC P24052;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Insulin-like growth factor I receptor precursor (BC 2.7.1.112).  
 GN IGFLR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95277910; PubMed=7758167;  
 RA Du J., Delafontaine P.;  
 RT "Inhibition of vascular smooth muscle cell growth through antisense  
 transcription of a rat insulin-like growth factor I receptor cDNA";  
 RL Circ. Res. 76:963-972(1995).  
 RN [2]  
 RP SEQUENCE OF 1-364 FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=90017496; PubMed=2477843;  
 RA Werner H., Woloschak M., Adamo M., Shen-Orr Z., Roberts C.T. Jr.,  
 RA Leroith D.;  
 RT "Developmental regulation of the rat insulin-like growth factor I  
 receptor gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7451-7455(1989).  
 RN [3]  
 RP SEQUENCE OF 913-1017 FROM N.A.  
 RX MEDLINE=92412145; PubMed=1530648;  
 RA Kurachi H., Jobo K., Ohta M., Kawasaki T., Itoh N.;  
 RT "A new member of the insulin receptor family, insulin  
 receptor-related receptor, is expressed preferentially in the  
 kidney";  
 RT Biochem. Biophys. Res. Commun. 187:934-939(1992).  
 CC -!- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)  
 WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A

CC TYROSINE-PROTEIN KINASE ACTIVITY.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -!- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide  
 CC bonds. The alpha chains contribute to the formation of the ligand-  
 CC binding domain, while the beta chain carries the kinase domain.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Belongs to the tyr family of protein kinases. Insulin  
 CC receptor subfamily.  
 CC -!- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L29232; AAA41392.1; -;  
 CC EMBL; M27293; AAA41384.1; -;  
 CC HSP; P06213; IIRK.  
 CC InterPro; IPR000494; EGFR\_L domain.  
 CC InterPro; IPR008957; FN\_III-like.  
 CC InterPro; IPR003961; FN\_III.  
 CC InterPro; IPR006211; Furin-like.  
 CC InterPro; IPR006212; Furin repeat.  
 CC InterPro; IPR009030; Growth factor receptor.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR002011; RecepttyrkinasII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00041; fn3; 2.  
 CC Pfam; PF00757; Furin-like; 1.  
 CC Pfam; PF00069; kinase; 1.  
 CC Pfam; PF01030; Recept\_L domain; 2.  
 CC PRINTS; PR0109; TYRKINASE.  
 CC PRODOM; PD000001; Prot\_kinase; 1.  
 CC SMART; SM00060; FN3; 3.  
 CC SMART; SM00261; FU; 1.  
 CC SMART; SM00219; Tyrc; 1.  
 CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 CC PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 CC PROSITE; PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
 CC PROSITE; PS50011; PROTEIN KINASE\_DOM; 1.  
 CC Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;  
 CC Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.  
 CC SIGNAL 1 30  
 CC CHAIN 31 741 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,  
 CC ALPHA-CHAIN.  
 CC CHAIN 742 1370 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,  
 CC BETA-CHAIN.  
 CC DOMAIN 742 936 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 937 960 POTENTIAL.  
 CC DOMAIN 961 1370 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 608 829 FIBRONECTIN TYPE-III 1.  
 CC DOMAIN 830 929 FIBRONECTIN TYPE-III 2.  
 CC DOMAIN 1000 1275 PROTEIN KINASE.  
 CC NP\_BIND 1006 1014 ATP (BY SIMILARITY).  
 CC BINDING 1034 1034 ATP (BY SIMILARITY).  
 CC ACT\_SITE 1136 1136 BY SIMILARITY.  
 CC DISULFID 215 224 BY SIMILARITY.  
 CC DISULFID 219 230 BY SIMILARITY.  
 CC DISULFID 231 239 BY SIMILARITY.  
 CC DISULFID 235 248 BY SIMILARITY.  
 CC DISULFID 251 260 BY SIMILARITY.  
 CC DISULFID 264 276 BY SIMILARITY.  
 CC DISULFID 282 303 BY SIMILARITY.  
 CC DISULFID 307 321 BY SIMILARITY.  
 CC DISULFID 324 328 BY SIMILARITY.  
 CC CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 641 641 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MOD RES 1166 1166 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CONFLICT 985 986 AD -> PY (IN REF. 3). CRC64;  
 SQ SEQUENCE 1370 AA; 155395 MM; A5946897A41CB145 CRC64;  
 Query Match 9.7%; Score 149; DB 1; Length 1370;  
 Best Local Similarity 28.7%; Pred. No. 0.0039;  
 Matches 47; Conservative 20; Mismatches 57; Indels 40; Gaps 11;  
 QY 34 HPWVGCGCGGATCSDYNGCLCKPRLFFVLRIKMKQIGVCLSSCPGYYGTRYPDIN 93  
 DB 232 HPE---CLGCHTPDINTTCVACHYY-----KGVCVPACPGTY--RP--- 271  
 QY 94 KCTKCKVDCITCFN-KNFCCTKCKSGFYHLGKCLDSCEGLEANN-HTMVCVSVHCEAS 151  
 DB 272 EGWRC-VDRDFCANIPNABSSDSGFGVHDEGCEQBCPSGFIRNSTQSMYCIP---CE-- 325  
 QY 152 EWSFPWPKMKGKTCGFKRGTRVTRDILQHPKAGK-----KGNL 191  
 DB 326 -----GPC---PRVCGDEBKTKTIDSVTAQMLQGCITLKGNL 361  
 RESULT 11  
 FSPO XENLA STANDARD; PRT; 803 AA.  
 AC P3547;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE F-spondin precursor.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93376785; PubMed=8367492;  
 RA Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;  
 RT "Ectopic neural expression of a floor plate marker in frog embryos  
 RT injected with the midline transcription factor Pintallavis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993)  
 CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory  
 CC neuron cells and the outgrowth of neurites in vitro. May  
 CC contribute to the growth and guidance of axons in both the spinal  
 CC cord and the PNS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.  
 CC -!- SIMILARITY: Contains 6 TSP type-1 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; L09123; AAA19105.1; -  
 CC PIR; A47723; A47723.

DR InterPro; IPR002861; Reeler.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF02014; Reeler; 1.  
 DR Pfam; PF00090; tsp\_1; 6.  
 DR SMART; SM00209; TSP1; 6.  
 DR PROSITE; PS00092; TSP1; 6.  
 KW Glycoprotein; Signal; Repeat; Cell adhesion.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 803 F-SPONDIN.  
 FT DOMAIN 436 489 TSP TYPE-1 1.  
 FT DOMAIN 495 549 TSP TYPE-1 2.  
 FT DOMAIN 552 605 TSP TYPE-1 3.  
 FT DOMAIN 608 662 TSP TYPE-1 4.  
 FT DOMAIN 664 717 TSP TYPE-1 5.  
 FT DOMAIN 750 802 TSP TYPE-1 6.  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 803 AA; 90702 MM; D3A54E329548ED9 CRC64;  
 Query Match 9.5%; Score 145.5; DB 1; Length 803;  
 Best Local Similarity 26.6%; Pred. No. 0.0041;  
 Matches 58; Conservative 23; Mismatches 84; Indels 53; Gaps 12;  
 QY 57 CKPRLFFVLE-----RIGMKQIGVCLSSCPGYYGTR-YPOINKCTCKV---- 100  
 DB 548 CEPSSCIVTVAEWERCSATCRMKKGRHMIKTPAD--GSMCKADTTEVEKCMPECH 605  
 QY 101 -----DCDTCFNKFTCKGSGFYHLGKCLDSCEGLEANN-----HTMBCV 143  
 DB 606 TIPCVLSPWSEWSDCVTCGKTRTRQR---MLKSPSELGDCNEBELKQVEKCMLEPCP 662  
 QY 144 SIHVCEASWSPKCKGKCTCGFKRGTRVTRDILQHPKAGKGNLCPPTSETRCIV 203  
 DB 663 --ISCLETWSYWSSEC---NKSG--KGHMIRMTMTWEPQFGA---VCPETVORKKC-- 711  
 QY 204 QRKCKSGRGGKGRKRRKRLKLNKRRKETSSTSSDSKG 241  
 DB 712 RLKCKQK-----SSGNRRHLKDKAREKSEKIKEDSDG 745  
 RESULT 12  
 NT2 RAT STANDARD; PRT; 2471 AA.  
 AC Q9QW10;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).  
 GN NOTCH2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93202015; PubMed=1295745;  
 RA Weinmaster G., Roberts V.J., Lemke G.;  
 RT "Notch2: a second mammalian Notch gene.";  
 RL Development 116:931-941(1992).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=21331789; PubMed=11438922;  
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;  
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple  
 RT functional roles for the Notch-DSL signaling system during brain  
 RT development.";  
 RL J. Comp. Neurol. 436:167-181(2001).  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
 CC Upon ligand activation through the released notch intracellular  
 CC domain (NICD) it forms a transcriptional activator complex with  
 CC RBP-J kappa and activates genes of the enhancer of split locus.



CC Affects the implementation of differentiation, proliferation and  
 CC apoptotic programs. May play an essential role in postimplantation  
 CC development, probably in some aspect of cell specification and/or  
 CC differentiation (By similarity).  
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 CC terminal fragment N(EGF) which are probably linked by disulfide  
 CC bonds (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 CC proteolytical processing NICD is translocated to the nucleus.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid  
 CC plexus in the brain. Expressed in postnatal central nervous system  
 CC (CNS) germinal zones and, in early postnatal life, within numerous  
 CC cells throughout the CNS. It is more highly localized to  
 CC ventricular germinal zones. Also found in the heart, liver and  
 CC kidney.  
 CC -1- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.  
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 CC which is proteolytically cleaved by a furin-like convertase in the  
 CC trans-Golgi network before it reaches the plasma membrane to yield  
 CC an active ligand-accessible form. Cleavage results in a C-  
 CC terminal fragment N(TM) and a N-terminal fragment N(EGF). Following  
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
 CC (TACE) to yield a membrane-associated intermediate fragment called  
 CC notch extracellular truncation (NEXT). This fragment is then  
 CC cleaved by presenilin dependent gamma-secretase to release a  
 CC notch-derived peptide containing the intracellular domain (NICD)  
 CC from the membrane (By similarity).  
 CC -1- PTM: Phosphorylated (By similarity).  
 CC -1- SIMILARITY: Belongs to the NOTCH family.  
 CC -1- SIMILARITY: Contains 35 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.  
 CC -1- SIMILARITY: Contains 6 ANK repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M93661; AAK1358.1; --  
 CC PIR: A49128; A49128.  
 CC HSP: P00743; ICCP.  
 CC InterPro: IPR002110; ANK.  
 CC InterPro: IPR000152; Asx\_hydroxyl\_S.  
 CC InterPro: IPR000742; EGF\_2.  
 CC InterPro: IPR001881; EGF\_Ca.  
 CC InterPro: IPR001438; EGF\_II.  
 CC InterPro: IPR005209; EGF\_like.  
 CC InterPro: IPR002049; Laminin\_EGF.  
 CC InterPro: IPR008297; Notch.  
 CC InterPro: IPR008000; Notch\_dom.  
 CC Pfam: PF00023; ank; 6.  
 CC Pfam: PF00008; EGF; 35.  
 CC Pfam: PF00066; notch; 2.  
 CC PIRSF: PIRSF002279; Notch; 1.  
 CC PRINTS: PR00010; EGFLOOD.  
 CC PRINTS: PR00011; EGFAMININ.  
 CC PRINTS: PR01452; NOTCH.  
 CC SMART: SM00248; ANK; 6.  
 CC SMART: SM00179; EGF\_CA; 24.  
 CC SMART: SM00004; NL; 2.  
 CC PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 CC PROSITE: PS50088; ANK\_REPEAT; 4.  
 CC PROSITE: PS00010; ASX\_HYDROXYL; 22.  
 CC PROSITE: PS00022; EGF\_1; 34.  
 CC PROSITE: PS01186; EGF\_2; 26.  
 CC PROSITE: PS50026; EGF\_3; 35.  
 CC PROSITE: PS01187; EGF\_CA; 22.  
 CC Receptor; Transcription regulation; Activator; Differentiation;  
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.

FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	2471	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT	CHAIN	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY
FT				SIMILARITY).
FT	CHAIN	1697	2471	NOTCH INTRACELLULAR DOMAIN (BY
FT				SIMILARITY).
FT	DOMAIN	26	1677	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1678	1698	POTENTIAL.
FT	DOMAIN	1699	2471	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	26	63	EGF-LIKE 1.
FT	DOMAIN	64	102	EGF-LIKE 2.
FT	DOMAIN	105	143	EGF-LIKE 3.
FT	DOMAIN	144	180	EGF-LIKE 4.
FT	DOMAIN	182	219	EGF-LIKE 5.
FT	DOMAIN	221	258	EGF-LIKE 6.
FT	DOMAIN	260	296	EGF-LIKE 7.
FT	DOMAIN	298	336	EGF-LIKE 8.
FT	DOMAIN	338	374	EGF-LIKE 9.
FT	DOMAIN	375	413	EGF-LIKE 10.
FT	DOMAIN	415	454	EGF-LIKE 11.
FT	DOMAIN	456	492	EGF-LIKE 12.
FT	DOMAIN	494	530	EGF-LIKE 13.
FT	DOMAIN	532	568	EGF-LIKE 14.
FT	DOMAIN	570	605	EGF-LIKE 15.
FT	DOMAIN	607	643	EGF-LIKE 16.
FT	DOMAIN	645	680	EGF-LIKE 17.
FT	DOMAIN	682	718	EGF-LIKE 18.
FT	DOMAIN	720	755	EGF-LIKE 19.
FT	DOMAIN	757	793	EGF-LIKE 20.
FT	DOMAIN	795	831	EGF-LIKE 21.
FT	DOMAIN	833	871	EGF-LIKE 22.
FT	DOMAIN	873	909	EGF-LIKE 23.
FT	DOMAIN	911	947	EGF-LIKE 24.
FT	DOMAIN	949	985	EGF-LIKE 25.
FT	DOMAIN	987	1023	EGF-LIKE 26.
FT	DOMAIN	1025	1061	EGF-LIKE 27.
FT	DOMAIN	1063	1099	EGF-LIKE 28.
FT	DOMAIN	1101	1147	EGF-LIKE 29.
FT	DOMAIN	1149	1185	EGF-LIKE 30.
FT	DOMAIN	1187	1223	EGF-LIKE 31.
FT	DOMAIN	1225	1262	EGF-LIKE 32.
FT	DOMAIN	1264	1302	EGF-LIKE 33.
FT	DOMAIN	1304	1343	EGF-LIKE 34.
FT	DOMAIN	1374	1412	EGF-LIKE 35.
FT	DOMAIN	1645	1648	POLY-ALA.
FT	DOMAIN	1994	1997	POLY-LEU.
FT	DOMAIN	2426	2429	POLY-SER.
FT	DOMAIN	2446	2451	POLY-GLY.
FT	REPEAT	1420	1456	LIN/NOTCH 1.
FT	REPEAT	1503	1535	LIN/NOTCH 2.
FT	REPEAT	1827	1871	ANK 1.
FT	REPEAT	1876	1905	ANK 2.
FT	REPEAT	1909	1939	ANK 3.
FT	REPEAT	1943	1972	ANK 4.
FT	REPEAT	1976	2005	ANK 5.
FT	REPEAT	2009	2038	ANK 6.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	90	BY SIMILARITY.
FT	DISULFID	92	101	BY SIMILARITY.
FT	DISULFID	109	121	BY SIMILARITY.
FT	DISULFID	115	131	BY SIMILARITY.
FT	DISULFID	133	142	BY SIMILARITY.
FT	DISULFID	148	159	BY SIMILARITY.
FT	DISULFID	153	168	BY SIMILARITY.
FT	DISULFID	170	179	BY SIMILARITY.
FT	DISULFID	186	198	BY SIMILARITY.
FT	DISULFID	192	207	BY SIMILARITY.
FT	DISULFID	209	218	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	230	246	BY SIMILARITY.

```

FT DISULFID 248 257 BY SIMILARITY.
FT DISULFID 264 275 BY SIMILARITY.
FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.

Query Match 9.4%; Score 144.5; DB 1; Length 2471;
Best Local Similarity 24.9%; Pred. No. 0.014;
Matches 55; Conservative 25; Mismatches 54; Indels 87; Gaps 14;

QY 30 QRRMHNVSQGGCGCATSDYNGCLSCRPFLPFVLERIGMKQIGVCLSCSPSGYGYTRY 89
DB 947 QTDMECLSEPCXNG-GTCSDYVNSYTC-----TCPPAGPHGVHC 984
QY 90 P-DINKCTKCKVDCDTCFKNFC-----TKKSGFYHLGK 124
DB 985 ENNIDECTE-----SSCFNGTCVGDINGSFCLCPVGTGPPCLHDINECSSNPLNSGT 1039
QY 125 CLD-----SCPEGLEANNHTMVCVSIH-CEASESWSPSCMKKGTGCPKRGTTETRV 176
DB 1040 CVDGLGTREYCTPLGYTGN-----CQTLVNLG-----SP-SPCKNG-TCQAQKA----- 1083
QY 177 RDILOHPSAKGNLCPTSETRTCTIVQRKCKSGKGRGKGG 217
DB 1084 -----RPRC-----LCPPGWDGAYCDVLANVSC-KAAALQNG 1113

RESULT 13
PSPO.RAT STANDARD; PRT; 807 AA.
AC P35446;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE F-spondin precursor.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic floor plate;
RX MEDLINE=92208952; PubMed=1555244;
RA Klar A., Baldaasare M., Jessell T.M.;
RT "F-spondin: a gene expressed at high levels in the floor plate
RT encodes a secreted protein that promotes neural cell adhesion and
RT neurite extension."
RL Cell 69:95-110(1992).
CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
CC neuron cells and the outgrowth of neurites in vitro. May
CC contribute to the growth and guidance of axons in both the spinal
CC cord and the PNS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
CC -!- SIMILARITY: Contains 6 TSP type-1 domains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR M88469; AAA41174.1; -
DR PIR; A38152; A38152.
DR InterPro; IPR002861; Reeler.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF02014; Reeler; 1.
DR Pfam; PF00990; TSP1; 6.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS0092; TSP1; 6.
KW Glycoprotein; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 28
FT CHAIN 29 807 F-SPONDIN.
FT DOMAIN 442 455 TSP TYPE-1 1.
FT DOMAIN 501 555 TSP TYPE-1 2.
FT DOMAIN 558 611 TSP TYPE-1 3.
FT DOMAIN 614 666 TSP TYPE-1 4.
FT DOMAIN 668 721 TSP TYPE-1 5.
FT DOMAIN 754 806 TSP TYPE-1 6.
FT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 681 681 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 807 AA; 90773 MW; 309525F9EAFEA89A CRC64;

Query Match 9.3%; Score 142.5; DB 1; Length 807;
Best Local Similarity 29.5%; Pred. No. 0.0066;
Matches 54; Conservative 16; Mismatches 60; Indels 53; Gaps 10;

QY 126 LDSCPEGLBANNHTM--ECVSIHVCEASWSPSCMKKGTGCPKRGTTETRVLDLQHP 183
DB 647 LGDCNEDLBOAEKMLPECP--IDCLSEWSQWSEC---NKSG--KGMIRTRTIQMEP 699
QY 184 SAKGKGNLCPTSETRTCTIVQRKCKSGK-----RGKKGRERKCKLNRK----- 229
DB 700 QFGGAP--CPET-----VQRKCKRARKLSPSIQKLRWRREARRSRSEQLRHBEDGE 750
QY 230 -----RKETSSSDSKGLSSIEPTDQENKRRQOQKRRARD-----KQKQSVSVS 276
DB 751 QPFGCEMRPTAWSECTKLGCGGIQ-----ERYVTYKPKSKSQFTSCDKKEIRAC 802
QY 277 TVH 279
DB 803 NVH 805

RESULT 14
IGIR.MOUSE STANDARD; PRT; 1373 AA.
AC Q60751; O70438; Q62123;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin-like growth factor I receptor precursor (EC 2.7.1.112).
GN IGIR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Navarro M., Garandel V., Barenton B., Bernardi H.;
RT "Cloning of cDNA for the mouse insulin-like growth factor I
RT receptor."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-329 FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;

```

RA Jun W., Liu Z., Alvares K., Kumar A., Wallner B.I., Kanwar Y.S.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1134-1203 FROM N.A.  
 RX MEDLINE=90152381; PubMed=2482828;  
 RA Wilks A.F., Kurban R.R., Hovens C.M., Ralph S.J.;  
 RT "The application of the polymerase chain reaction to cloning members  
 of the protein tyrosine kinase family.";  
 RL Gene 85:67-74(1989).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)  
 WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A  
 TYROSINE-PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 tyrosine phosphate.  
 CC -1- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide  
 bonds. The alpha chains contribute to the formation of the ligand-  
 binding domain, while the beta chain carries the kinase domain.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin  
 receptor subfamily.  
 CC -1- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF056187; AAC12782.1; -;  
 DR EMBL; U00182; AAC52123.1; -;  
 DR EMBL; M33422; AAA40013.1; -;  
 DR PIR; A48805; A48805.  
 DR HSP; P06213; IIRK.  
 DR MGD; MGI:96433; Igflr.  
 DR GO; GO:0009887; P: organogenesis; IMP.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR006211; Purin-like.  
 DR InterPro; IPR006212; Purin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002011; Recepttyr\_kinII.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR008266; Tyr\_pkinase\_AS.  
 DR Pfam; PF00041; fn3; 2.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recept\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00060; FN3; 3.  
 DR SMART; SM00261; FU; 1.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00111; PROTEIN KINASE\_DOM; 1.  
 DR PROSITE; PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
 DR Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;  
 Glycoprotein; ATP-binding; phosphorylation; Repeat; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 741 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,  
 ALPHA-CHAIN  
 FT CHAIN 742 1373 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,  
 BETA-CHAIN.  
 FT DOMAIN 742 936 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 937 960 CYTOPLASMIC (BY SIMILARITY).  
 FT DOMAIN 961 1373 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 608 829 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 830 929 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 1000 1276 PROTEIN KINASE.

FT NP\_BIND 1006 1014 ATP (BY SIMILARITY).  
 FT BINDING 1034 1034 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1137 1137 BY SIMILARITY.  
 FT DISULFID 215 224 BY SIMILARITY.  
 FT DISULFID 219 230 BY SIMILARITY.  
 FT DISULFID 231 239 BY SIMILARITY.  
 FT DISULFID 235 248 BY SIMILARITY.  
 FT DISULFID 251 260 BY SIMILARITY.  
 FT DISULFID 264 276 BY SIMILARITY.  
 FT DISULFID 282 303 BY SIMILARITY.  
 FT DISULFID 307 321 BY SIMILARITY.  
 FT DISULFID 324 328 BY SIMILARITY.  
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 535 535 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 641 641 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 757 757 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 765 765 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT MOD\_RES 1167 1167 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CONFLICT 58 59 FL -> LV (IN REF. 2).  
 FT CONFLICT 260 260 C -> S (IN REF. 2).  
 FT CONFLICT 301 301 D -> G (IN REF. 2).  
 FT CONFLICT 306 306 E -> V (IN REF. 2).  
 FT CONFLICT 324 324 C -> S (IN REF. 2).  
 FT CONFLICT 1134 1134 V -> I (IN REF. 3).  
 FT CONFLICT 1145 1145 V -> D (IN REF. 3).  
 FT CONFLICT 1202 1202 V -> I (IN REF. 3).  
 SQ SEQUENCE 1373 AA; 155787 MW; 58E3B72EF101E379 CRC64;  
 Query Match 9.3%; Score 142; DB 1; Length 1373;  
 Best Local Similarity 28.0%; Pred. No. 0.012;  
 Matches 46; Conservative 20; Mismatches 59; Indels 40; Gaps 11;  
 Qy 34 HPNVSGGGGCGATCSYNGCLSCFPLFFVLEIGMKQIGVCLSSCPGSGYGYTPDIN 93  
 Db 232 HPE---CLGSCHTPDNTTCVCRHYIY-----KGVCPACPGTY--RF---- 271  
 Qy 94 KCTKCKVDCDTCEN-KNFTCKSKGFYLHKGKLDSCPEGLEANN-HTMBCVSVIHCERAS 151  
 Db 272 EGWRC-VDRDFCANIPNABSSDSOGFVHDECHQBCPSGFIKNSQSMYCLP---CB--- 325  
 Qy 152 EWSPWSPQMKKGTGCFRGRTGTETVRDILQHPKAG-----KGNL 191  
 Db 326 -----GFC---PKVCGDEBKTKTIDSVTSAGMLQGCTILKGNL 361  
 RESULT 15  
 NTC2 MOUSE  
 ID NTC2 MOUSE STANDARD; PRT; 2470 AA.  
 AC O35516; Q06008; Q60941.  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DB Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch  
 B).  
 GN NOTCH2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6; TISSUE=Thymus;

RA Hamada Y., Higuchi M., Teujimoto Y.;  
RT "Complete amino acid sequence and multiform transcripts encoded by a  
RT single copy of mouse Notch2 gene.";  
RN Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [2].  
RC SEQUENCE OF 316-1518 FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
RX MEDLINE=93178563; PubMed=8440332;  
RA Lardelli M., Lendahl U.;  
RT "Moch and Notch B-two mouse Notch homologues coexpressed in a  
RT wide variety of tissues.";  
RL Exp. Cell Res. 204:364-372(1993).  
RN [3].  
RP SEQUENCE OF 1765-2153 FROM N.A.  
RX MEDLINE=97075110; PubMed=8917536;  
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,  
RA Martin D.I.;  
RT "Inhibition of granulocytic differentiation by mNotch1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).  
RN [4].  
RP FUNCTION.  
RX MEDLINE=99396706; PubMed=10393120;  
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,  
RA Teujimoto Y.;  
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
RT embryonic lethality.";  
RL Development 126:3415-3424 (1999).  
RN [5].  
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.  
RX MEDLINE=95333893; PubMed=7609614;  
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Teujimoto Y.;  
RT "Differential expression of Notch1 and Notch2 in developing and adult  
RT mouse brain.";  
RL Brain Res. Mol. Brain Res. 29:263-272(1995).  
RN [6].  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RX MEDLINE=21523956; PubMed=11518718;  
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (NL-4) undergo presenilin-dependent  
RT proteolysis.";  
RL J. Biol. Chem. 276:40268-40273(2001).  
RN [7].  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RX MEDLINE=21374376; PubMed=11459941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).  
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation.  
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=035516-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=035516-2; Sequence=VSP\_001405;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
CC neuroepithelia, somites, optic vesicles and branchial arches, but  
CC not heart.  
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,

CC the postnatal ependymal cells, and the choroid plexus throughout  
CC embryonic and postnatal development.  
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TMP-alpha converting enzyme  
CC (TACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC -1- PTM: Phosphorylated.  
CC -1- SIMILARITY: Belongs to the NOTCH family.  
CC -1- SIMILARITY: Contains 35 EGF-like domains.  
CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.  
CC -1- SIMILARITY: Contains 6 ANK repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D32210; BAA22094.1; -;  
CC EMBL: X68279; CAA48340.1; -;  
CC EMBL: U31881; AAC52924.1; -;  
CC PIR: A49175; A49175.  
CC HSP: P16109; IFSB.  
CC MGD; MGI:97364; Notch2.  
CC GO; GO:0005887; C: integral to plasma membrane; IC.  
CC GO; GO:0005157; P: protein binding; IPI.  
CC GO; GO:0002011; P: morphogenesis of an epithelial sheet; IMP.  
CC GO; GO:0007219; P: N signaling pathway; IC.  
CC InterPro; IPR002110; ANK.  
CC InterPro; IPR000152; Ax\_hydroxyl\_S.  
CC InterPro; IPR000742; EGF\_2.  
CC InterPro; IPR001881; EGF\_Ca.  
CC InterPro; IPR001438; EGF\_II.  
CC InterPro; IPR006209; EGF\_like.  
CC InterPro; IPR002049; Laminin\_EGF.  
CC InterPro; IPR008297; Notch.  
CC InterPro; IPR000800; Notch\_dom.  
CC Pfam; PF00023; ank; 6.  
CC Pfam; PF00066; EGF; 34.  
CC PIRSP; PIRSP002279; Notch; 1.  
CC PRINTS; PR00010; EGFLOOD.  
CC PRINTS; PR00011; EGF\_LAMININ.  
CC PRINTS; PR01452; NOTCH.  
CC SMART; SM00248; ANK; 6.  
CC SMART; SM00179; EGF\_Ca; 23.  
CC SMART; SM00004; NL\_3.  
CC PROSITE; PS0297; ANK\_REPEAT; 1.  
CC PROSITE; PS00088; ANK\_REPEAT; 4.  
CC PROSITE; PS00010; ASX\_HYDROXYL; 22.  
CC PROSITE; PS00022; EGF\_1; 33.  
CC PROSITE; PS01186; EGF\_2; 27.  
CC PROSITE; PS00026; EGF\_3; 35.  
CC PROSITE; PS01187; EGF\_Ca; 22.  
KW Receptor; Transcription regulation; Activator; Differentiation;  
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;  
KW Alternative splicing.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.  
FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.  
FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1678 1698 POTENTIAL.

FT	DOMAIN	1699	2470	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	26	63	EGF-LIKE 1.
FT	DOMAIN	64	102	EGF-LIKE 2.
FT	DOMAIN	105	143	EGF-LIKE 3.
FT	DOMAIN	144	180	EGF-LIKE 4.
FT	DOMAIN	182	219	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	221	256	EGF-LIKE 6 (INCOMPLETE).
FT	DOMAIN	258	294	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	296	334	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	336	372	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	373	411	EGF-LIKE 10.
FT	DOMAIN	413	452	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	454	490	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	492	528	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	530	566	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	568	603	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	605	641	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	643	678	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	680	716	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	718	753	EGF-LIKE 19.
FT	DOMAIN	755	791	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	793	829	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	831	869	EGF-LIKE 22.
FT	DOMAIN	871	907	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	909	945	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	947	983	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	985	1021	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1023	1059	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1061	1097	EGF-LIKE 28.
FT	DOMAIN	1099	1145	EGF-LIKE 29.
FT	DOMAIN	1147	1183	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1185	1221	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1223	1260	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1262	1300	EGF-LIKE 33.
FT	DOMAIN	1302	1345	EGF-LIKE 34.
FT	DOMAIN	1372	1410	EGF-LIKE 35.
FT	REPEAT	1418	1454	LIN/NOTCH 1.
FT	REPEAT	1501	1533	LIN/NOTCH 2.
FT	REPEAT	1825	1869	ANK 1.

Query Match 9.2%; Score 140.5; DB 1; Length 2470;  
Best Local Similarity 24.4%; Pred. No. 0.026;  
Matches 54; Conservative 24; Mismatches 56; Indels 87; Gaps 14;

QY	30	QRRMHPNVSGCGGCGCATCDYNGCLSCPKRLPFVLERIGHKQIGVCLSSCPSGGYGYTRY	89
Db	945	QTDNNECLSEPCKNG-GTCSDYNSVTC-----TCPAGFHGVHC	982
QY	90	P-DINKCTKCKVDCDTCFNKFC-----TKCKSGFYHLGK	124
Db	983	ENNIDCTB-----SSCFNGGTCVDGINSFSCICPVGFTGPFCLHDINECSNPLNAGT	1037
QY	125	CLDS-----CPGLEANNHMECVSIVH-CEASWSWSPQMKKTKCGFKRGTTTRV	176
Db	1038	CVDGLGYRCICPLGYTKY-----CQTLVNLCSR-----SPCKNG-TC-----VQEKA	1081
QY	177	RDILQHPNAGKGNLCPPTSETCTICVQRKCKSGERKKG	217
Db	1082	R-----PHC-----LCPPGWDGAYCDLVNSC-KAALQKG	1111

Search completed: June 29, 2004, 17:00:17  
Job time : 9.46481 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:55:24 ; Search time 35.5522 Seconds  
(without alignments)  
2476.067 Million cell updates/sec

Title: US-09-894-912A-32

Perfect score: 1535

Sequence: 1 MHLRLISCFIILNFMFYIG.....QQRRARDKQKSVSVSTVH 279

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1315.5	85.7	272	4 Q9BXV4	Q9bxv4 homo sapien
2	1273	82.9	292	4 Q96K87	Q96k87 homo sapien
3	1212	79.0	217	11 Q9CSB2	Q9cab2 mus musculus
4	1107	72.1	224	11 Q8BVW2	Q8bvw2 mus musculus
5	632	41.2	265	11 Q9Z1J2	Q9z1j2 mus musculus
6	616.5	40.2	236	4 Q8N7L5	Q8n7l5 homo sapien
7	565.5	36.8	243	11 Q8BFU0	Q8bfu0 mus musculus
8	450	29.3	224	4 Q9UGB2	Q9ugb2 homo sapien
9	449.5	29.3	176	4 Q8N6X6	Q8n6x6 homo sapien
10	419.5	27.3	152	11 Q7TPX3	Q7tpx3 mus musculus
11	283.5	18.5	138	11 Q8BU73	Q8bu73 mus musculus
12	185.5	12.1	1299	5 Q26489	Q26489 spodoptera
13	174	11.3	1376	5 Q8S252	Q8s252 drosophila
14	171	11.1	296	11 Q35171	Q35171 mus musculus
15	171	11.1	826	11 Q8CF22	Q8cf22 mus musculus
16	171	11.1	932	11 Q62030	Q62030 mus musculus

17	165	10.8	915	11	Q91VK0	Q91vk0 mus musculus
18	165.5	10.8	913	13	Q8AY18	Q8ay18 rana esculle
19	162.5	10.6	1101	5	Q964D2	Q964d2 entamoeba h
20	161.5	10.5	803	13	Q42114	Q42114 brachydanio
21	160.5	10.5	503	5	Q9U018	Q9u018 giardia lam
22	160.5	10.5	1074	5	Q964D1	Q964d1 entamoeba h
23	160	10.4	808	13	Q42113	Q42113 brachydanio
24	154	10.0	913	4	Q968P4	Q968p4 homo sapien
25	154	10.0	4010	11	Q80T14	Q80t14 mus musculus
26	151.5	9.9	3869	5	Q86PQ3	Q86pq3 cryptospori
27	149	9.7	1371	11	Q9QVW4	Q9qvww rattus sp.
28	147.5	9.6	548	5	Q9GQ45	Q9gq45 giardia lam
29	144	9.4	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
30	143	9.3	4007	4	Q86XX4	Q86xx4 homo sapien
31	142	9.3	898	5	Q76822	Q76822 branchiosco
32	141.5	9.2	402	11	Q8K2Q8	Q8k2q8 mus musculus
33	141.5	9.2	807	11	Q8VCC9	Q8vcc9 mus musculus
34	140.5	9.2	213	11	Q991E4	Q991e4 mus musculus
35	140.5	9.2	214	11	Q8R0Y1	Q8r0y1 mus musculus
36	140.5	9.2	660	5	Q23832	Q23832 cryptospori
37	139.5	9.1	807	4	Q9HCB6	Q9hcb6 homo sapien
38	139	9.1	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
39	138.5	9.0	724	4	Q94862	Q94862 homo sapien
40	138.5	9.0	807	4	Q8NCD7	Q8ncd7 homo sapien
41	137	8.9	440	5	O18003	O18003 caenorhabdi
42	137	8.9	441	4	Q96JW7	Q96jw7 homo sapien
43	137	8.9	651	4	Q86UZ4	Q86uz4 homo sapien
44	136.5	8.9	642	5	Q8MPM6	Q8mpm6 giardia lam
45	136.5	8.9	739	5	Q9GS24	Q9gs24 giardia lam

#### ALIGNMENTS

#### RESULT 1

Q9BXV4 ID Q9BXV4 PRELIMINARY; PRT; 272 AA.  
AC Q9BXV4, 2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mao Y., Xie Y., Zhou Z., Zhao W., Wang S., Huang Y., Wang S.,  
Tang R., Chen X., Wu C.,  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF251057; AAH22367.1; -.  
DR EMBL; BC022367; AAH22367.1; -.  
DR Genew; HGNC:20866; THSD2.  
DR InterPro; IPR006212; Furin repeat.  
DR InterPro; IPR009030; Grow\_fac\_recep.  
DR InterPro; IPR00884; TSP1.  
DR Pfam; PFO0090; tsp\_1; 1.  
DR SMART; SM00261; FU; 2.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS50092; TSP1; 1.  
SQ SEQUENCE 272 AA; 30928 MW; CACAC6B7E781189 CRC64;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;

Best Local Similarity 87.1%; Pred. No. 1.8e-110;

Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFIILNFMFYIGSQRRARDKQKSVSVSTVHNGCLSKRP 60

```

Db      1  MLRLISLWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCSYNGCLSCKPR 60
QY      61  LFPVLERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCFNKNFCTKCKSGFY 120
Db      61  LFPALERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCFNKNFCTKCKSGFY 120
QY      121  HLKGLDSCPEGLRANNTMBCVSIHVCEASEWSPWSPCKKGTCTGKRGTTETVRDIL 180
Db      121  HLKGLDNCPEGLRANNTMBCVSIHVCEASEWSPWSPCKKGTCTGKRGTTETVRDIL 180
QY      181  QHPKAGKGNLCPPTSETTCTVQIRKCKSGRGRKRKRLKLNKBERKETSSSDSK 240
Db      181  QHPSA--KGNLCPTTETKCTVQIRKCKSGRGRKRKRLKLNKBERKETSSSDSK 240
QY      241  GLESSIETPDQENKEROQKQKARDKQKSVSVSTVH 279
Db      237  SLESSKEIPEQRENK--QOKKRVQDK-QKSVSVSTVH 272

RESULT 2
Q96K87 PRELIMINARY; PRT; 292 AA.
AC Q96K87;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14440.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027346; BAB55051.1; -.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp.1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3D4A6F8 CRC64;

Query Match 82.9%; Score 1273; DB 4; Length 292;
Best Local Similarity 84.4%; Pred. No. 1.3e-106;
Matches 233; Conservative 14; Mismatches 23; Indels 6; Gaps 3;

```

```

QY      1  MLRLISLWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCSYNGCLSCKPR 60
Db      237  SLESSKEIPEQRENK--QOKKRVQDKQKSGIEVT 270

RESULT 3
Q9CSB2 PRELIMINARY; PRT; 217 AA.
AC Q9CSB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 2810459H04RIK protein (Pragmat).
GN THSD2 OR 2810459H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gibsi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK013366; BAB28811.1; -.
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp.1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
FT NON_TER 217 AA; 24304 MW; 0DCF938E9B3BF7 CRC64;
SQ SEQUENCE 217 AA; 24304 MW; 0DCF938E9B3BF7 CRC64;

Query Match 79.0%; Score 1212; DB 11; Length 217;
Best Local Similarity 98.6%; Pred. No. 3e-101;
Matches 216; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

```

```

QY      1  MLRLISLWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCSYNGCLSCKPR 60
Db      1  MLRLISLWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCSYNGCLSCKPR 60
QY      61  LFPVLERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCFNKNFCTKCKSGFY 120
Db      61  LFPALERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCFNKNFCTKCKSGFY 120
QY      121  HLKGLDSCPEGLRANNTMBCVSIHVCEASEWSPWSPCKKGTCTGKRGTTETVRDIL 180
Db      121  HLKGLDNCPEGLRANNTMBCVSIHVCEASEWSPWSPCKKGTCTGKRGTTETVRDIL 180
QY      181  QHPKAGKGNLCPPTSETTCTVQIRKCKSGRGRKRKRLKLNKBERKETSSSDSK 240
Db      181  QHPSA--KGNLCPTTETKCTVQIRKCKSGRGRKRKRLKLNKBERKETSSSDSK 240
QY      241  GLESSIETPDQENKEROQKQKARDKQKSVSVSTVH 279
Db      237  SLESSKEIPEQRENK--QOKKRVQDK-QKSVSVSTVH 272

```



## RESULT 4

Q8BVW2 PRELIMINARY; PRT; 224 AA.  
 AC Q8BVW2; TREMBLrel. 23, Created  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Thrombospondin homolog.  
 GN THSD2 OR 2810459H04RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 .OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK076308; BAC36296.1; -.  
 DR MGD; MGI:1920030; Thsd2.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR SMART; SM00261; FU; 2.  
 SQ SEQUENCE 224 AA; 25398 MW; BCL3B083497CFEE3 CRC64;

Query Match 72.1%; Score 1107; DB 11; Length 224;  
 Best Local Similarity 88.4%; Pred. No. 8.7e-92;  
 Matches 198; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

Qy 1 MHLRLISCFPIILNFMVEYIGSNASRGRRORRPMHPNVUSQGGCGATCSDYNGCLSKPRL 60  
 Db 1 MHLRLISCFPIILNFMVEYIGSNASRGRRORRPMHPNVUSQGGCGATCSDYNGCLSKPRL 60  
 Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKTKCKVDCDTCFNKNFCTKCKSGPYL 120  
 Db 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKTKCKVDCDTCFNKNFCTKCKSGPYL 120  
 Qy 121 HLGKCLDSCPGLRANNTMCVSIHVCEASESWSPCMKKGKTCGFKGTETRVRDIL 180  
 Db 121 HLGKCLDSCPGLRANNTMCVSIHVCEASESWSPCMKKGKTCGFKGTETRVRDIL 180  
 Qy 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGRGGKGRKRRKXK 224  
 Db 181 QHPSA--KGNVPVPPNQDRDNLYSTKGVFKGRARKGKRETKK 222

## RESULT 5

Q92132 PRELIMINARY; PRT; 265 AA.  
 AC Q92132;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Thrombospondin type 1 domain.  
 GN RSPONDIN OR R-SPONDIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 .OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Kanata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;  
 RA "R-spondin, a novel thrombospondin type 1 domain gene, expressed in  
 RT the dorsal neural tube";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016768; BAA75640.1; -.  
 DR MGD; MGI:2183426; Rspndin.

DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR SMART; SM00261; FU; 2.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 SQ SEQUENCE 265 AA; 29331 MW; FFEB8964743F5963 CRC64;

Query Match 41.2%; Score 632; DB 11; Length 265;

Best Local Similarity 43.7%; Pred. No. 5.8e-49;  
 Matches 114; Conservative 42; Mismatches 91; Indels 14; Gaps 4;

Qy 3 LRLISCFPIILNFMVEYIGSNASRGRRORRPMHPNVUSQGGCGATCSDYNGCLSKPRL 61  
 Db 1 MRLGLCVVALVLSWTHIAVSGRIGKGRQRISAEBSQACAKGCLCSVNGCLCSPKL 60  
 Qy 62 FFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKTKCKVD-CDTFKNKFNCTKCKSGPYL 120  
 Db 61 FILLERNDIRGVCLPSCPFGYDARNPDMNKCIKCLIEHCEAFSHNFTKQBEALYL 120  
 Qy 121 HLGKCLDSCPGLRANNTMCVSIHVCEASESWSPCMKKGKTCGFKGTETRVRDIL 180  
 Db 121 HKGRCPACPEGSTAANSTMEGSPAQCESESWSPMGPCSKRKLCPGPKSGSEETRVRL 180  
 Qy 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGRGGKGRKRRKRLKLNKZERKKTSSSDSK 240  
 Db 181 HAPG--GDHTTCSDTKTRKCTVTRTPCPBGQKRGKGGQGRRENANRHPARKNSKEPRS- 237  
 Qy 241 GLESSIETPDQENKEROQQ 261  
 Db 238 -----NSRRHKGGQQPQ 249

## RESULT 6

Q8N7L5 PRELIMINARY; PRT; 236 AA.  
 AC Q8N7L5;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ40906.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 .OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Oshima A., Takahashi-Fujii A., Tanase T., Inose N., Takeuchi K.,  
 RA Arita M., Mueashino K., Yuuki H., Hara H., Sugiyama T., Irie K.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,  
 RA Sugano S., Nagahori K., Masuho Y., Nagai K., Ito G.,  
 RA "NEDO human cDNA sequencing project";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK098225; BAC05263.1; -.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR SMART; SM00261; FU; 2.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 236 AA; 25942 MW; 8D03803127EC5678 CRC64;

Query Match 40.2%; Score 616.5; DB 4; Length 236;  
 Best Local Similarity 47.3%; Pred. No. 1.3e-47;  
 Matches 112; Conservative 32; Mismatches 72; Indels 21; Gaps 4;

QY 32 RMHPNVSGCGGATCSDYNGCLSCPKPLFFVLERIGMKQIGVCLSSCPGYYGTRYPD 91  
 Db 4 RVSAEGSQAKAGCCLSEVNGCLSKSPKFIILLERNDIRQVGVCLSCPFGVFDARNPD 63  
 QY 92 INKTKCKVD-COTCFNNKFTCKGSGFYHLGKCLDSCPEGLBANNHTMECVSIHVCEA 150  
 Db 64 MNKCIKIEHCACFHNFTCKEGLYLHGRCYPACPEGSSAANGTMECSPPAQCEV 123  
 QY 151 SEWSPWSPCKKGTCTCFKGTETVRVDILQHPGAKGNLCPTSTRTCTIVORCKSK 210  
 Db 124 SEWSPWSPCKKGTCTCFKGTETVRVDILQHPGAKGNLCPTSTRTCTIVORCKSK 210  
 QY 211 GEGKKG---RRKRKLANKERKTSSSSDSKGLESIETTPDQGNKEROQOQKR 263  
 Db 182 GQYKGGGQGRNANRNLARKSKKAGACG-----RRKGGQOQOQOQ 224

## RESULT 7

Q8BFUO PRELIMINARY; PRT; 243 AA.  
 AC Q8BFUO  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Hypothetical thrombospondin type I repeat.  
 GN 2610028P08RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Rodentia; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Eye, and Hippocampus;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK049891; BAC33974.1; -  
 DR EMBL; AK087485; BAC39893.1; -  
 DR MGD; MGI:1922667; 2610028P08RIK.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR00884; TSPI.  
 DR SMART; SM00261; FU; 2.  
 DR SMART; SM00209; TSPI; 1.  
 DR PROSITE; PS50092; TSPI; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 243 AA; 28275 MW; ED76A08D61012ED7 CRC64;

Query Match 36.8%; Score 565.5; DB 11; Length 243;  
 Best Local Similarity 45.8%; Pred. No. 5.2e-43;  
 Matches 110; Conservative 36; Mismatches 83; Indels 11; Gaps 7;  
 QY 1 MLRLISCFPIILNFMBEYIGSQNARGRORRHPNVSGCGGATCSDYNGCLSCPKR 60  
 Db 1 MRCLFSFALLIINCNDYSQC-GNRWRNKA-SYVSNFICKGLSCSKDNGSCRCQK 58  
 QY 61 LFFVLERIGMKQIGVCLSSCPGYYGTRYPDKTKCKVD-CDYCFNKNFTCKKSGFY 119  
 Db 59 LFFFLRREGWRQYGECLHSCFSGYGYGHRAFDNRCARCIENCDSCFSDCTCKKGVY 118  
 QY 120 LHLGKCLDSCPEGLBANNHTMECVSIHVCEASWSPWSPCKKGTCTCFKGTETVRDI 179  
 Db 119 LHRGRCFDECFDGFAPLDETMECVS--GCEVGHSEWGTCSRNRNCTGFWGLETRTQI 176  
 QY 180 LQHPGAKGNL-CPTSTRTCTIVORCKSKGERGKGRERKRLNKR--ERKETSSS 236  
 Db 177 VKAPA---KDTIPCTIAESRCMMNRHCPGGKRTPKAKEKRNKGRKRLIERAQSHS 233  
 RESULT 8

Q9UGB2 PRELIMINARY; PRT; 224 AA.  
 AC Q9UGB2  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE D7824P16.3 (Novel protein similar to mouse thrombospondin type 1  
 DE domain protein R-spondin) (Fragment).  
 GN D7824P16.3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Blakey S.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL050325; CAB65783.3; -  
 DR Genew; HGNC:16175; C20orf182.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR00884; TSPI.  
 DR SMART; SM00261; FU; 2.  
 DR SMART; SM00209; TSPI; 1.  
 DR PROSITE; PS50092; TSPI; 1.  
 FT NON TER 224 224  
 SQ SEQUENCE 224 AA; 25042 MW; 97D26AD34CDBF812 CRC64;  
 Query Match 29.3%; Score 450; DB 4; Length 224;  
 Best Local Similarity 42.1%; Pred. No. 1.2e-32;  
 Matches 88; Conservative 33; Mismatches 70; Indels 18; Gaps 7;  
 QY 28 RRQRHPNVSGCGGATCSDYNGCLSCPKPLFFVLERIGMKQIGVCLSSCPGYYGT 87  
 Db 22 RRKXQVGTGLGNCCT-GCIICSENGCSTCOQLFLFIRREGIKQYKGLHDCPPGYFGI 80  
 QY 88 RYDPINKTKCKVDCTCFNKNFTCKGSGFYHLGKCLDSCPEGLBANNHTMECVSIHV 147  
 Db 81 RGQEVNRCKKGATCSCFSDQFCIRCKRQFYLYKGLCLCTCPPTGLAHQNTRECOG--B 138  
 QY 148 CEASWSPWSPCKKGTCTCFKGTETVRVDI--LQHPGAKGNLCPTSTRTCTIVOR 205  
 Db 139 CELPGWGSPTCTHNGKTCGSAWGLSESRVREAGRACHEAA----TQVLSESRKCPQR 194  
 QY 206 KCKSKGERG---KKGRE----RKRKLNK 227  
 Db 195 P--CPGERSPGQKGRKRRPRKDKLDR 221  
 RESULT 9  
 Q8N6X6 PRELIMINARY; PRT; 176 AA.  
 AC Q8N6X6  
 DT 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Similar to putative.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC027938; AAT27938.1; -  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR00884; TSPI.  
 DR SMART; SM00261; FU; 1.  
 DR SMART; SM00209; TSPI; 1.  
 DR PROSITE; PS50092; TSPI; 1.

Db 59 LFFFLRREGMRQYGECLHSCPSGYGHRAPDMNRCARCIENCDSCFSKDFCTCKVGEY 111

Qy 120 LHLGKCLDSCPEGLAANNHTMECVSIHVHCEASWS 154  
 ||:|||||:|||||  
 Db 119 LHRGRCFDECFDGFAPLDETMECVE--GCEVGHS 151  
 ||:|||||:|||||

RESULT 11

Q8BJ73 PRELIMINARY; PRT; 138 AA.

AC Q8BJ73;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE DJ824P16.3.  
 DE A930029K19RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RX MEDLINE=22354683; PubMed=12456851;  
 RA THE FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs";  
 RL Nature 420:563-573 (2002).  
 DR ENBL; AK020304; BAC25643.1; -  
 DR MGD; MGI:1924467; A930029K19RIK.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR SMART; SM00261; FU; 2.  
 SQ SEQUENCE 138 AA; 15172 MW; F8FD7D949279D5DD CRC64;

Query Match 18.5%; Score 283.5; DB 11; Length 138;  
 Best Local Similarity 42.9%; Pred. No. 6.9e-18;  
 Matches 45; Conservative 23; Mismatches 36; Indels 1; Gaps 1

Qy 28 RQRRRHPNVSQCGSCAFCSDYNGCLCKPFLFVLERIGMKQIGVCLSSCPSGYGT 87  
 ||:::|:||||:||||:||||:||||:||||:||||:||||:||||:|  
 Db 22 RRRQAGTGGGCT--GCVCSENGSGTCCQRLFLIRREGIRQYKGVHDCPLGFFGI 80  
 ||:::|:||||:||||:||||:||||:||||:||||:||||:||||:|

Qy 88 RYPDINKTKCKVDCTCFNKNFCTCKSGFYHLGLKCLDSCPEG 132  
 ||:|||||:||||:||||:||||:||||:||||:||||:|  
 Db 81 RGQEAENCKKGATCSCFSDFCIRKRRPHLYKGKCLPSCP TG 125  
 ||:|||||:||||:||||:||||:||||:||||:||||:|

RESULT 12

Q26489 PRELIMINARY; PRT; 1299 AA.

AC Q26489;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Endoprotease furin.  
 GN FURIN  
 OS Spodoptera frugiperda (Fall armyworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
 OC Noctuidae; Amphipyrinae; Spodoptera.  
 OC NCBI\_TaxID=7108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RC Clepik M., Klenk H.;  
 RA "Cloning and functional characterization of FURIN from Spodoptera  
 frugiperda (sf9) cells";  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
 DR ENBL; Z68888; CAA93116.1; -.



[illegible]

165 FDSLVKKSGCHTCTCTCGPSRBBCLHCAKSFHFQDMKVCAPACGSGFYPEEMPGLPHKV 224 Db

159 CMKKGKTC---GFKRG-----TETVRVDILQHPSAKGNKLCPPPTSETTCIVORK 206 Qy

225 CARCENCLSCGGSRNCRCKAGFTOLGTSCTINHTCSNADETFCBNWKSRLNC---ERK 282 Db

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:09 ; Search time 49.1845 Seconds  
(without alignments)  
1562.545 Million cell updates/sec

Title: US-09-894-912A-34  
Perfect score: 1510  
Sequence: 1 MHLRLISWLFILFMFYIG.....QQKKRVQDKSVSVSTVH 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1510	100.0	272	4 AAB99220	Aab99220 Human thr
2	1510	100.0	272	4 AAM78328	Aam78328 Human pro
3	1510	100.0	272	4 AAE13168	Aae13168 Human ste
4	1510	100.0	272	4 AAE13150	Aae13150 Human ste
5	1510	100.0	272	6 ABR62112	AbR62112 Human clo
6	1510	100.0	272	6 ABO44432	AbO44432 Human ste
7	1510	100.0	272	6 ABO44413	AbO44413 Human ste
8	1505	99.7	273	4 AAE13151	Aae13151 Human ste
9	1505	99.7	273	6 ABO44414	AbO44414 Human ste
10	1477	97.8	265	4 AAE13163	Aae13163 Human ste
11	1477	97.8	265	6 ABO44427	AbO44427 Human sec
12	1477	97.8	292	2 AAW85607	Aaw85607 Secreted
13	1477	97.8	292	4 AAE13170	Aae13170 Human SCR
14	1477	97.8	292	5 AWP61846	Awp61846 Human pol
15	1477	97.8	292	6 ABR62114	AbR62114 Human sec
16	1477	97.8	292	6 ABO44434	AbO44434 Human SCR
17	1457	96.5	292	4 AAB93875	Aab93875 Human pro
18	1400	92.7	251	4 AAE13153	Aae13153 Human nat
19	1400	92.7	251	6 ABO44417	AbO44417 Human ste
20	1315.5	87.1	279	4 AAE13167	Aae13167 Mouse ste
21	1315.5	87.1	279	6 ABO44431	AbO44431 Mouse ste
22	1183	78.3	239	6 ABUS2396	Abu52396 Human GPC
23	1111	73.6	195	6 ABUS2398	Abu52398 Human GPC
24	1107	73.3	195	6 ABUS2397	Abu52397 Human GPC
25	903	59.8	160	4 ABB11374	Abb11374 Human sec

26	903	59.8	160	4 AAM79312	Aam79312 Human pro
27	903	59.8	160	4 AAE13149	Aae13149 Human ste
28	903	59.8	160	6 ABO44415	AbO44415 Human ste
29	656	43.4	263	6 ABR62108	AbR62108 Secreted
30	656	43.4	263	6 ABR62115	AbR62115 Secreted
31	656	43.4	263	6 ABR58489	AbR58489 Human sec
32	656	43.4	263	7 ADB07919	AdB07919 Novel pro
33	648.5	42.9	243	6 ABR62110	AbR62110 Secreted
34	644	42.6	265	6 ABR62113	AbR62113 Mouse thr
35	638	42.3	229	4 AAE13162	Aae13162 Mouse thr
36	638	42.3	229	6 ABO44426	AbO44426 Mouse thr
37	589.5	39.0	243	5 AAE37115	Aae37115 Human sec
38	582.5	38.6	243	5 ABR76508	AbR76508 DNA encod
39	582.5	38.6	243	6 ABR62106	AbR62106 Secreted
40	578.5	38.3	243	6 AAE36166	Aae36166 Mouse SCR
41	578	38.3	250	6 ABR62101	AbR62101 Secreted
42	545	36.1	222	6 ABR62107	AbR62107 Secreted
43	540.5	35.8	229	6 ABR62102	AbR62102 Secreted
44	472	31.3	234	7 ADB76146	AdB76146 Novel hum
45	276	18.3	46	4 AAE13155	Aae13155 Human SCR

## ALIGNMENTS

RESULT 1  
AAB99220  
ID AAB99220 standard; protein; 272 AA.  
XX  
AC AAB99220;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Human thrombospondin-30.  
XX  
KW Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;  
KW malignant tumor; haemopathy; HIV infection; immunological disease;  
KW inflammation disease.  
XX  
OS Homo sapiens.  
XX  
FN WO200140294-A1.  
XX  
PD 07-JUN-2001.  
XX  
PF 27-NOV-2000; 2000WO-CN000494.  
XX  
PR 29-NOV-1999; 99CN-00124148.  
XX  
PA (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.  
XX  
PI Mao Y, Xie Y;  
XX  
DR MPI; 2001-397948/42.  
DR N-PSDB; AAB45131.  
XX  
PT Human thrombospondin-30 and polynucleotide is useful in diagnosis and  
PT treatment of, e.g., malignant tumor, hemopathy, HIV infection,  
PT immunological diseases and various inflammatory diseases.  
XX  
PS Claim 1; Page 27-28; 33pp; Chinese.  
XX  
CC The present sequence is the protein sequence for human thrombospondin-30.  
CC Thrombospondin-30 protein and coding sequence are useful in the diagnosis  
CC and treatment of malignant tumor, haemopathy, HIV infection,  
CC immunological diseases and various inflammation diseases. In addition  
CC thrombospondin-30 protein may be used for screening mimics, agonists,  
CC antagonists or inhibitors, or for use in peptide fingerprinting  
CC identification. The thrombospondin-30 coding sequence may be used as  
CC primers for nucleic acid amplification reaction or as probes for  
CC hybridisation reaction, or in producing gene chips or microarrays  
SQ Sequence 272 AA;

```

Query Match      100.0%; Score 1510; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 1e-110;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISWLFILNFMFYIGSONASRGRRRRHHPNVISQGGCGCATCSDYNGCLSCCKPR 60
DB 1 MHLRLISWLFILNFMFYIGSONASRGRRRRHHPNVISQGGCGCATCSDYNGCLSCCKPR 60

QY 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYDPDKCTCKADCDTCFNNKFNCTCKSGFY 120
DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYDPDKCTCKADCDTCFNNKFNCTCKSGFY 120

QY 121 HLKGLCLNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180
DB 121 HLKGLCLNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180

QY 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGKGRKRRKPKNGSKKEAIPDSKSLES 240
DB 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGKGRKRRKPKNGSKKEAIPDSKSLES 240

QY 241 SKETPEQRENKQKQKRVQDKQKSVSVTVH 272
DB 241 SKETPEQRENKQKQKRVQDKQKSVSVTVH 272

RESULT 2
AAW78328
ID AAW78328 standard; protein; 272 AA.
AC AAW78328;
XX
XX 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 990.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX W0200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 27-APR-2000; 2000US-00560875.
XX
XX 20-JUN-2000; 2000US-00598075.
XX
XX 19-JUL-2000; 2000US-00620325.
XX
XX 01-SEP-2000; 2000US-00654936.
XX
XX 15-SEP-2000; 2000US-00663561.
XX
XX 20-OCT-2000; 2000US-00693325.
XX
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAK51461.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 20; Page 3214-3215; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

```

```

CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
XX Sequence 272 AA;
XX
XX Query Match      100.0%; Score 1510; DB 4; Length 272;
XX Best Local Similarity 100.0%; Pred. No. 1e-110;
XX Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MHLRLISWLFILNFMFYIGSONASRGRRRRHHPNVISQGGCGCATCSDYNGCLSCCKPR 60
XX DB 1 MHLRLISWLFILNFMFYIGSONASRGRRRRHHPNVISQGGCGCATCSDYNGCLSCCKPR 60
XX
XX QY 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYDPDKCTCKADCDTCFNNKFNCTCKSGFY 120
XX DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYDPDKCTCKADCDTCFNNKFNCTCKSGFY 120
XX
XX QY 121 HLKGLCLNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180
XX DB 121 HLKGLCLNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180
XX
XX QY 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGKGRKRRKPKNGSKKEAIPDSKSLES 240
XX DB 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGKGRKRRKPKNGSKKEAIPDSKSLES 240
XX
XX QY 241 SKETPEQRENKQKQKRVQDKQKSVSVTVH 272
XX DB 241 SKETPEQRENKQKQKRVQDKQKSVSVTVH 272
XX
XX RESULT 3
XX AAE13168
XX ID AAE13168 standard; protein; 272 AA.
XX
XX AC AAE13168;
XX
XX XX 28-JAN-2002 (first entry)
XX
XX XX Human stem cell growth factor-like protein #4.
XX
XX Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;
XX neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;
XX vasotropic; virucide; dermatological; tranquillisier; cerebroprotective;
XX osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
XX duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
XX acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
XX Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
XX adrenal white matter degeneration; anaemia; neurodegenerative disease;
XX Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
XX severe combined immunodeficiency; immune disorder; autoimmune disorder;
XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
XX autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
XX supporting factor for the proliferation of stem cell.
XX
XX Homo sapiens.
XX
XX OS
XX Key Location/Qualifiers
XX Peptide 1..21
XX Protein /label= Signal_peptide
XX /note= "Human mature stem cell growth factor-like
XX protein"
XX

```



PN WO200177169-A2.  
 XX 18-OCT-2001.  
 PD  
 XX  
 PF 05-APR-2001; 2001WO-US011208.  
 XX  
 XX 05-APR-2000; 2000US-00543774.  
 PR 28-JUN-2000; 2000US-0215733P.  
 PR 09-JAN-2001; 2001US-00757562.  
 PR 05-FEB-2001; 2001US-0266614P.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA (KIRI ) KIRIN BEER KK.  
 PA  
 XX  
 XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;  
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
 PI  
 XX WPI; 2001-657166/75.  
 DR N-PSDB; AAD21728.  
 DR  
 XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.  
 PT  
 XX Claim 28; Page 226-227; 232pp; English.  
 XX  
 CC The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumours. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopoenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency,  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 CC inflammation. Sequences of the invention are also useful in gene therapy.  
 CC The present sequence is stem cell growth factor-like protein from human  
 XX  
 XX Sequence 272 AA;  
 SQ  
 Query Match 100.0%; Score 1510; DB 4; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1e-110;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MHLRLISLWPIILNFMFYISQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKCPR 60  
 DB 1 MHLRLISLWPIILNFMFYISQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKCPR 60  
 QY 61 LFFALERIGHKQIGVCLSSCPSGYGYTRYPDINKCTCKADCDTCFNKFNCTCKSGPYL 120  
 DB 61 LFFALERIGHKQIGVCLSSCPSGYGYTRYPDINKCTCKADCDTCFNKFNCTCKSGPYL 120  
 QY 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTCKGKTCTCFKGTETRVREII 180  
 DB 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTCKGKTCTCFKGTETRVREII 180  
 QY 181 QHPSAKGNLCPTNTRKCTVQRKCKQGRGKKGRKRRKKNKGNKSKRAIPDSKSLS 240  
 DB 181 QHPSAKGNLCPTNTRKCTVQRKCKQGRGKKGRKRRKKNKGNKSKRAIPDSKSLS 240

QY 241 SKBIPGRENKQKQKRVQDKQKSVSVSTVH 272  
 DB 241 SKBIPGRENKQKQKRVQDKQKSVSVSTVH 272  
 RESULT 4  
 AAE13150  
 ID AAE13150 standard; protein; 272 AA.  
 XX  
 AC AAE13150;  
 XX  
 XX 28-JAN-2002 (first entry)  
 DT  
 XX Human stem cell growth factor-like protein #2.  
 DE  
 XX Human; stem cell growth factor-like protein; antiinflammatory; nootropic;  
 KW neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;  
 KW vasotrophic; virucide; dermatological; tranquiliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopoenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 KW supporting factor for the proliferation of stem cell.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT Protein /label= Signal\_peptide  
 FT 22..272  
 FT /note= "Human mature stem cell growth factor-like  
 FT protein"  
 FT  
 XX WO200177169-A2.  
 XX  
 XX 18-OCT-2001.  
 XX  
 XX 05-APR-2001; 2001WO-US011208.  
 XX  
 XX 05-APR-2000; 2000US-00543774.  
 PR 28-JUN-2000; 2000US-0215733P.  
 PR 09-JAN-2001; 2001US-00757562.  
 PR 05-FEB-2001; 2001US-0266614P.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA (KIRI ) KIRIN BEER KK.  
 PA  
 XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;  
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
 PI  
 XX WPI; 2001-657166/75.  
 DR N-PSDB; AAD21728.  
 DR  
 XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.  
 PT  
 XX Claim 28; Page 211-212; 232pp; English.  
 XX  
 CC The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow

transplantation or cord blood transplantation for treating a variety of diseases such as immunodeficiency syndrome, chronic granulomatous disease, duplicated immunodeficiency syndrome, agammaglobulinaemia, Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS), thalassemia, haemolytic anaemia due to enzyme defect, congenital anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal storage diseases such as mucopolysaccharidosis, adrenal white matter degeneration, a variety of cancer and tumours. Proteins of the invention are useful for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, thrombocytopaenia, immune deficiencies and disorders such as severe combined immunodeficiency (SCID) and autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary inflammation. Sequences of the invention are also useful in gene therapy. The present sequence is stem cell growth factor-like protein from human

XX Sequence 272 AA;

Query Match 100.0%; Score 1510; DB 4; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1e-110;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISMLFIILNFMFYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKR 60  
 |||||  
 DB 1 MHLRLISMLFIILNFMFYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKR 60  
 |||||

QY 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKCTKCKADCTCFNKNFCTCKSGFYL 120  
 |||||  
 DB 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKCTKCKADCTCFNKNFCTCKSGFYL 120  
 |||||

QY 121 HLGKCLDNCPEGLEANNHTMECVSIHVCSVSEWNPSPCTKGGTCTGFRGTETRVREII 180  
 |||||  
 DB 121 HLGKCLDNCPEGLEANNHTMECVSIHVCSVSEWNPSPCTKGGTCTGFRGTETRVREII 180  
 |||||

QY 181 QHPSAKGNLCPPNTRKCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240  
 |||||  
 DB 181 QHPSAKGNLCPPNTRKCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240  
 |||||

QY 241 SKEIPEQRNKKQKKRKVDKQKSVSVTVH 272  
 |||||  
 DB 241 SKEIPEQRNKKQKKRKVDKQKSVSVTVH 272  
 |||||

# RESULT 5

ID ABR62112 standard; protein; 272 AA.

XX AC ABR62112;

XX 18-AUG-2003 (first entry)

DE Human clone 1 thrombospondin protein #23.

XX Human; secreted; stem cell growth factor; cytostatic; haemostatic;  
 KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;  
 KW degenerative disease; Alzheimer's disease; food supplement;  
 KW immunological disorder; thrombospondin.

XX Homo sapiens.

XX WO2003029405-A2.

XX 10-APR-2003.

XX 30-AUG-2002; 2002WO-US027746.

XX 30-AUG-2001; 2001US-0316368P.

XX 10-DEC-2001; 2001US-0339739P.

XX 19-APR-2002; 2002US-00125852.

XX (HYSB-) HYSBQ INC.

XX Tang YT;

XX

DR WPI; 2003-381616/36.

XX New stem cell growth factor-like polypeptides and polynucleotides, useful  
 PT for treating e.g. leukemia, hemophilia and degenerative diseases like  
 PT Alzheimer's disease, and for inducing immune response.

XX Disclosure; Fig 1; 15pp; English.

XX The invention relates to new stem cell growth factor-like polypeptides  
 CC and polynucleotides. The stem cell growth factor-like polypeptides and  
 CC polynucleotides are useful for inducing differentiation of embryonic and  
 CC adult stem cells to give rise to different cell types, for treating e.g.  
 CC leukaemia, haemophilia and degenerative diseases like Alzheimer's  
 CC disease. They are also useful for generating new tissues and organs that  
 CC may aid patients in need of transplanted tissues. The polynucleotides are  
 CC useful as hybridisation probes, oligomers or primers for PCR, for  
 CC chromosome and gene mapping, in recombinantly producing protein, in  
 CC generating antisense DNA or RNA, in diagnostics as expressed sequence  
 CC tags for identifying expressed genes, and for inducing immune response.  
 CC The polypeptides are useful for generating antibodies that specifically  
 CC bind the polypeptide, as molecular weight markers, and as a food  
 CC supplement (e.g. protein or amino acid supplement, and as a carbon,  
 CC nitrogen or carbohydrate source). Compositions comprising the  
 CC polypeptides or polynucleotides are useful for the diagnosis, treatment  
 CC or prevention of cancers, and other immunological disorders. The current  
 CC sequence represents a human clone 1 thrombospondin protein

XX Sequence 272 AA;

Query Match 100.0%; Score 1510; DB 6; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1e-110;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISMLFIILNFMFYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKR 60  
 |||||  
 DB 1 MHLRLISMLFIILNFMFYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKR 60  
 |||||

QY 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKCTKCKADCTCFNKNFCTCKSGFYL 120  
 |||||  
 DB 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKCTKCKADCTCFNKNFCTCKSGFYL 120  
 |||||

QY 121 HLGKCLDNCPEGLEANNHTMECVSIHVCSVSEWNPSPCTKGGTCTGFRGTETRVREII 180  
 |||||  
 DB 121 HLGKCLDNCPEGLEANNHTMECVSIHVCSVSEWNPSPCTKGGTCTGFRGTETRVREII 180  
 |||||

QY 181 QHPSAKGNLCPPNTRKCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240  
 |||||  
 DB 181 QHPSAKGNLCPPNTRKCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240  
 |||||

QY 241 SKEIPEQRNKKQKKRKVDKQKSVSVTVH 272  
 |||||  
 DB 241 SKEIPEQRNKKQKKRKVDKQKSVSVTVH 272  
 |||||

RESULT 6

ABO44432

ID ABO44432 standard; protein; 272 AA.

XX AC ABO44432;

XX 30-SEP-2003 (first entry)

DE Human stem cell growth factor-like protein, SCR 1 #4.

XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
 KW immunostimulant; vulnery; haematopoietic stem cell; gene therapy;  
 KW supporting factor for proliferation of stem cells; wound healing;  
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;  
 KW bone marrow transplantation; cord blood transplantation;  
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;  
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;  
 KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;

KW epithelial cell growth; ovarian follicle development; nerve cell growth;  
 KW cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1. 21

XX Protein /label= Signal\_peptide

XX 22. 272

XX /notes= "Mature stem cell growth factor-like protein. This  
 XX protein is specifically claimed in claim 9"

XX US2003044792-A1.

XX 06-MAR-2003.

XX 28-JUN-2001; 2001US-00894912.

XX 28-JUN-2000; 2000US-0215733P.

XX 05-FEB-2001; 2001US-0266614P.

XX 05-APR-2001; 2001US-0282397P.

XX (TANG/) TANG Y T.

XX (LABA/) LABAT I.

XX (DRMA/) DRMANAC R T.

XX (MIZE/) MIZE N.

XX (NISH/) NISHIKAWA M.

XX (CHAO/) CHAO C.

XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

XX WPI; 2003-625403/59.

XX N-PSDB; ACH04328.

XX Novel isolated polypeptide having stem cell growth factor activity,  
 XX useful for promoting wound healing, and as a medicine to proliferate or  
 XX support human hematopoietic stem cells or human hematopoietic progenitor  
 XX cells.

XX Claim 23; Page 82; 96pp; English.

XX The invention relates to an isolated stem cell growth factor-like  
 CC polypeptide (referred as supporting factor for proliferation of stem  
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or  
 CC fragment, analogue, variant or derivative, that retains stem cell growth  
 CC factor activity. Also included are an isolated polynucleotide encoding  
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or  
 CC derivative, that retains stem cell growth factor activity, or the  
 CC complement of the polynucleotide), an (expression) vector comprising the  
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the  
 CC SCR-1 polynucleotide in operative association with a regulatory sequence  
 CC that controls expression of the polynucleotide in the host cell,  
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an  
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an  
 CC activity to support proliferation or survival of hematopoietic stem cell  
 CC or hematopoietic progenitor cell, with a proviso that C-terminal aa  
 CC sequence does not comprise the aa sequence appearing as ABO44433), an  
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and  
 CC lacking any 10 consecutive aa from ABO44430, an isolated polypeptide  
 CC with stem cell growth factor activity having at least an aa sequence  
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1  
 CC polypeptide to maintain survival of or promote proliferation of a stem  
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array  
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1  
 CC polynucleotide attached to a surface, a stromal cell genetically  
 CC engineered to express the SCR-1 polypeptide to support proliferation or  
 CC survival of a stem cell or germ cell, and an implant comprising a cell  
 CC genetically engineered to express the SCR-1 polypeptide to support  
 CC proliferation or survival of a stem cell or germ cell. The SCR-1  
 CC polypeptide is useful for identifying a compound that binds to the SCR-1  
 CC polypeptide and for maintaining survival of or promoting proliferation of  
 CC a stem cell, a germ cell, a hematopoietic stem cell or a hematopoietic  
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound

CC healing. The human hematopoietic stem cell or human hematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of hematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia  
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's  
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a Human SCR-1  
 CC protein

XX Sequence 272 AA;

Query Match 100.0%; Score 1510; DB 6; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1e-110;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMGYIGSQNASRGRQRMRHPNVSGCGGCGCATCDYNGCLSKPR 60

Db 1 MHLRLISWLFILNFMGYIGSQNASRGRQRMRHPNVSGCGGCGCATCDYNGCLSKPR 60

Qy 61 LFPALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTKADCDTCFNKNFCTCKSGPYL 120

Db 61 LFPALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTKADCDTCFNKNFCTCKSGPYL 120

Qy 121 HLKGLDNCPEGLEANNHTMECVSIHVCESEWNPSPCTKKGKTCGFKRGTTETVREII 180

Db 121 HLKGLDNCPEGLEANNHTMECVSIHVCESEWNPSPCTKKGKTCGFKRGTTETVREII 180

Qy 181 QHPSAKGNLCPPNTNETKCTVQRKKCKGKGRKGRKPKPNKGSKEAIPDSKSL 240

Db 181 QHPSAKGNLCPPNTNETKCTVQRKKCKGKGRKGRKPKPNKGSKEAIPDSKSL 240

Qy 241 SKETPEORENKKQKKKQKQKSVSVTVH 272

Db 241 SKETPEORENKKQKKKQKQKSVSVTVH 272

RESULT 7

ABO44413

ID ABO44413 standard; protein; 272 AA.

XX ABO44413;

XX 30-SEP-2003 (first entry)

XX Human stem cell growth factor-like protein, SCR 1 #1.

XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
 KW immunostimulant; vulnery; hematopoietic stem cell; gene therapy;  
 KW supporting factor for proliferation of stem cells; wound healing;  
 KW hematopoietic progenitor cell; stromal cell; AIDS; thalassemia;  
 KW bone marrow transplantation; cord blood transplantation;  
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;  
 KW agammaglobulinemia; Wiskott-Aldrich syndrome; haemolytic anaemia;  
 KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;  
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;  
 KW cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.

XX US2003044792-A1.

XX 06-MAR-2003.

XX 28-JUN-2001; 2001US-00894912.

XX 28-JUN-2000; 2000US-0215733P.

XX 05-FEB-2001; 2001US-0266614P.

PR 05-APR-2001; 2001US-0282397P.  
 XX (TANG/) TANG Y T.  
 PA (LABA/) LABAT I.  
 PA (DRMA/) DRMANAC R T.  
 PA (MIZE/) MIZE N.  
 PA (NISH/) NISHIKAWA M.  
 PA (CHAO/) CHAO C.  
 XX  
 PI Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;  
 DR WPI; 2003-625403/59.  
 DR N-PSDB; ACH04323, ACH04324.  
 XX  
 PT Novel isolated polypeptide having stem cell growth factor activity,  
 PT useful for promoting wound healing, and as a medicine to proliferate or  
 PT support human hematopoietic stem cells or human hematopoietic progenitor  
 PT cells.  
 XX  
 PS Claim 23; Fig 3; 96pp; English.  
 CC The invention relates to an isolated stem cell growth factor-like  
 CC polypeptide (referred as supporting factor for proliferation of stem  
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or  
 CC fragment, analogue, variant or derivative, that retains stem cell growth  
 CC factor activity. Also included are an isolated polynucleotide encoding  
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or  
 CC derivative, that retains stem cell growth factor activity, or the  
 CC complement of the polynucleotide), an (expression) vector comprising the  
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the  
 CC SCR-1 polynucleotide in operative association with a regulatory sequence  
 CC that controls expression of the polynucleotide in the host cell,  
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an  
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an  
 CC activity to support proliferation or survival of hematopoietic stem cell  
 CC or hematopoietic progenitor cell, with a proviso that C-terminal aa  
 CC sequence does not comprise the aa sequence appearing as ABO44433), an  
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and  
 CC lacking any 10 consecutive aas from ABO44430, an isolated polypeptide  
 CC with stem cell growth factor activity having at least an aa sequence  
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1  
 CC polypeptide to maintain survival of or promote proliferation of a stem  
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array  
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1  
 CC polynucleotide attached to a surface, a stromal cell genetically  
 CC engineered to express the SCR-1 polypeptide to support proliferation or  
 CC survival of a stem cell or germ cell and an implant comprising a cell  
 CC genetically engineered to express the SCR-1 polypeptide to support  
 CC proliferation or survival of a stem cell or germ cell. The SCR-1  
 CC polypeptide is useful for identifying a compound that binds to the SCR-1  
 CC polypeptide and for maintaining survival of or promoting proliferation of  
 CC a stem cell, a germ cell, a hematopoietic stem cell or a hematopoietic  
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound  
 CC healing. The human hematopoietic stem cell or human hematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of hematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia,  
 CC due to enzyme defect, congenital anaemia such as sickle cell disease,  
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a Human SCR-1  
 CC protein  
 XX  
 SQ Sequence 272 AA;  
 Query Match 100.0%; Score 1510; DB 6; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1e-110;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISMLPIILNFMFYIGSONASRGRORRHPNVNVSQCGGCATCSDYNGCLSKPR 60  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1 MHLRLISMLPIILNFMFYIGSONASRGRORRHPNVNVSQCGGCATCSDYNGCLSKPR 60  
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYTRYPDINKCTCKKADCDTCFNRNPTCKKSGFYL 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 61 LFPALERIGMKQIGVCLSSCPGSGYTRYPDINKCTCKKADCDTCFNRNPTCKKSGFYL 120  
 QY 121 HLGKCLDNCPEGLEANNHTMBCVSIHVCEVSEKNPMSPTCKGKTGFKGTETRVREII 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 121 HLGKCLDNCPEGLEANNHTMBCVSIHVCEVSEKNPMSPTCKGKTGFKGTETRVREII 180  
 QY 181 QHPSAKGNLCPPNTRKCTVORKKCKGGRKKRERKRNKESKAI PDSKSLSES 240  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 181 QHPSAKGNLCPPNTRKCTVORKKCKGGRKKRERKRNKESKAI PDSKSLSES 240  
 QY 241 SKEIPEQRENKQKKRKVKQKSVSVSTVH 272  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 241 SKEIPEQRENKQKKRKVKQKSVSVSTVH 272

RESULT 8  
 AAE13151  
 ID AAE13151 standard; protein; 273 AA.  
 XX AAE13151;  
 AC AAE13151;  
 XX  
 DT 28-JAN-2002 (first entry)  
 XX  
 DE Human stem cell growth factor-like protein #3.  
 XX  
 KW Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;  
 KW neuroprotective; vulnerary; cytostatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 KW supporting factor for the proliferation of stem cell.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200177169-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 05-APR-2001; 2001WO-US011208.  
 XX  
 PR 05-APR-2000; 2000US-00543774.  
 PR 28-JUN-2000; 2000US-0215733P.  
 PR 09-JAN-2001; 2001US-00757562.  
 PR 05-FEB-2001; 2001US-0266614P.  
 XX  
 XX (HYSB-) HYSQ INC.  
 PA (KIRI ) KIRIN BEER KK.  
 XX  
 XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;  
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;  
 XX WPI; 2001-657166/75.  
 DR N-PSDB; AAD21725.  
 DR  
 XX  
 PT Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.  
 XX  
 PS Claim 28; Page 214-215; 232pp; English.

XX The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumours. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 CC inflammation. Sequences of the invention are also useful in gene therapy.  
 CC The present sequence is stem cell growth factor-like protein from human  
 XX Sequence 273 AA;

Query Match 99.7%; Score 1505; DB 4; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-110;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HRLISLWLFILNPMVEYIGSNASRGRRRRMHPNVSGCGGCGATGSDYNGCLSCPKRL 61  
 DB 3 HRLISLWLFILNPMVEYIGSNASRGRRRRMHPNVSGCGGCGATGSDYNGCLSCPKRL 62  
 QY 62 FFALERIGMKQIGVCLSSCPGGYGTGTPDINKCTCKACDCTCFNKNFTCKCKSGYLH 121  
 DB 63 FFALERIGMKQIGVCLSSCPGGYGTGTPDINKCTCKACDCTCFNKNFTCKCKSGYLH 122  
 QY 122 LGKCLDNCPLGANNHMTVCVSIHCVSEFWNPSPCTKKGKTCGFKGTETVRVRIIQ 181  
 DB 123 LGKCLDNCPLGANNHMTVCVSIHCVSEFWNPSPCTKKGKTCGFKGTETVRVRIIQ 182  
 QY 182 HPSAKGNLCPTTNETRKTCTVQRKKCKGKRGKGRKRRKPKNGESKGAIPDSKLSLESS 241  
 DB 183 HPSAKGNLCPTTNETRKTCTVQRKKCKGKRGKGRKRRKPKNGESKGAIPDSKLSLESS 242  
 QY 242 KEIPEQRENKQKKRKVKQDKSVSVSTVH 272  
 DB 243 KEIPEQRENKQKKRKVKQDKSVSVSTVH 273

RESULT 9  
 ABO44414  
 ID ABO44414 standard; protein; 273 AA.  
 XX AC ABO44414;  
 XX AC ABO44414;  
 XX 30-SEP-2003 (first entry)  
 XX Human stem cell growth factor-like protein, SCR 1 #2.  
 XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
 XX immunostimulant; vulnary; haematopoietic stem cell; gene therapy;  
 XX supporting factor for proliferation of stem cells; wound healing;  
 XX haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;  
 XX bone marrow transplantation; cord blood transplantation;  
 XX chronic granulomatous disease; duplicated immunodeficiency syndrome;  
 XX agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;  
 XX congenital anaemia; sickle cell; Gaucher's disease; morphogenesis;  
 XX epithelial cell growth; ovarian follicle development; nerve cell growth;  
 XX cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.  
 OS US2003044792-A1.  
 XX 06-MAR-2003.  
 XX 28-JUN-2001; 2001US-00894912.  
 XX 28-JUN-2000; 2000US-0215733P.  
 PR 05-FEB-2001; 2001US-0266614P.  
 PR 05-APR-2001; 2001US-0282397P.  
 XX (TANG/) TANG Y T.  
 PA (LABA/) LABAT I.  
 PA (DRMA/) DRMANAC R T.  
 PA (MIZE/) MIZE N.  
 PA (NISH/) NISHIKAWA M.  
 PA (CHAO/) CHAO C.  
 XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;  
 WIPI; 2003-625403/59.  
 DR N-PSDB; ACH04325.  
 XX Novel isolated polypeptide having stem cell growth factor activity,  
 PT useful for promoting wound healing, and as a medicine to proliferate or  
 PT support human hematopoietic stem cells or human hematopoietic progenitor  
 PT cells.  
 XX Claim 23; Page 72-73; 96pp; English.  
 CC The invention relates to an isolated stem cell growth factor-like  
 CC polypeptide (referred as supporting factor for proliferation of stem  
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or  
 CC fragment, analogue, variant or derivative, that retains stem cell growth  
 CC factor activity. Also included are an isolated polynucleotide encoding  
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or  
 CC derivative, that retains stem cell growth factor activity, or the  
 CC complement of the polynucleotide), an (expression) vector comprising the  
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the  
 CC SCR-1 polynucleotide in operative association with a regulatory sequence  
 CC that controls expression of the polynucleotide in the host cell,  
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an  
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an  
 CC activity to support proliferation or survival of haematopoietic stem cell  
 CC or haematopoietic progenitor cell, with a proviso that C-terminal aa  
 CC sequence does not comprise the aa sequence appearing as ABO44433), an  
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and  
 CC lacking any 10 consecutive aa from ABO44430, an isolated polypeptide  
 CC with stem cell growth factor activity having at least an aa sequence  
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1  
 CC polypeptide to maintain survival of or promote proliferation of a stem  
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array  
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1  
 CC polynucleotide attached to a surface, a stromal cell genetically  
 CC engineered to express the SCR-1 polypeptide to support proliferation or  
 CC survival of a stem cell or germ cell and an implant comprising a cell  
 CC genetically engineered to express the SCR-1 polypeptide. The SCR-1  
 CC polypeptide is useful for identifying a compound that binds to the SCR-1  
 CC polypeptide and for maintaining survival of or promoting proliferation of  
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic  
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound  
 CC healing. The human haematopoietic stem cell or human haematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of haematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia  
 CC due to enzyme defect, congenital anaemia such as sickle cell, Gaucher's  
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and

CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a Human SCR-1  
 CC protein  
 XX  
 SQ Sequence 273 AA;

Query Match 99.7%; Score 1505; DB 6; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-110;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HLRILSWLFIILNFMVEYIGSNASGRQRMRHPNVSGCGCATCDYNGCLSKPRL 61  
 Db 3 HLRILSWLFIILNFMVEYIGSNASGRQRMRHPNVSGCGCATCDYNGCLSKPRL 62  
 Qy 62 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYLH 121  
 Db 63 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYLH 122  
 Qy 122 LGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKGGTKCGPGRGTETVRREII 181  
 Db 123 LGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKGGTKCGPGRGTETVRREII 182  
 Qy 182 HPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLESS 241  
 Db 183 HPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLESS 242  
 Qy 242 KEIPEQRENKQKKRKVKQDKQKSVSVTVH 272  
 Db 243 KEIPEQRENKQKKRKVKQDKQKSVSVTVH 273

RESULT 10  
 AAE13163  
 ID AAE13163 standard; protein; 265 AA.

AC AAE13163;  
 XX  
 XX 28-JAN-2002 (first entry)  
 XX Human secreted protein from clone DA228\_6.

XX Human; stem cell growth factor-like protein; antiinflammatory; neutropic;  
 KW neuroprotective; vulnerable; cytostatic; anticonvulsant; immunostimulant;  
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;  
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;  
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;  
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;  
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;  
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;  
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;  
 KW severe combined immunodeficiency; immune disorder; autoimmune disease;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;  
 KW supporting factor for the proliferation of stem cell; secreted protein.

XX Homo sapiens.  
 XX W0200177169-A2.  
 XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.  
 XX 05-APR-2000; 2000US-00543774.  
 XX 28-JUN-2000; 2000US-0215733P.  
 XX 09-JAN-2001; 2001US-00757562.  
 XX 05-FEB-2001; 2001US-0266614P.

XX (HYSB-) HYSEQ INC.  
 XX (KIRI) KIRIN BEER KK.

XX

PI Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;  
 PI Sacho-Crain B, Dickson M, Mize NK, Nishikawa M;  
 XX WPI; 2001-657166/75.

XX Novel stem cell growth factor like polypeptides and polynucleotides for  
 PT identifying modulators useful for treating diseases such as Alzheimer's  
 PT disease, cancer, rheumatoid arthritis, osteoporosis.

XX Disclosure; Fig 3; 232pp; English.

XX The patent discloses novel stem cell growth factor-like proteins and  
 CC polynucleotides encoding them. Proteins of the invention are also known  
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem  
 CC cell growth factor-like proteins are useful for supporting proliferation  
 CC or survival of a stem cell or germ cell which is preferably primordial  
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem  
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent  
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth  
 CC factor-like proteins can replace as a graft for the bone marrow  
 CC transplantation or cord blood transplantation for treating a variety of  
 CC diseases such as immunodeficiency syndrome, chronic granulomatous  
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),  
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia  
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage  
 CC diseases such as mucopolysaccharidosis, adrenal white matter  
 CC degeneration, a variety of cancer and tumours. Proteins of the invention  
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune  
 CC deficiencies and disorders such as severe combined immunodeficiency  
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic  
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary  
 CC inflammation. Sequences of the invention are also useful in gene therapy.  
 CC The present sequence is human secreted protein from clone DA228\_6  
 XX  
 SQ Sequence 265 AA;

Query Match 97.8%; Score 1477; DB 4; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-108;  
 Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELRLISWLFILNFMVEYIGSNASGRQRMRHPNVSGCGCATCDYNGCLSKPRL 60  
 Db 1 MELRLISWLFILNFMVEYIGSNASGRQRMRHPNVSGCGCATCDYNGCLSKPRL 60  
 Qy 61 LFFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYL 120  
 Db 61 LFFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYL 120  
 Qy 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKGGTKCGPGRGTETVRREII 180  
 Db 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKGGTKCGPGRGTETVRREII 180  
 Qy 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLES 240  
 Db 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLES 240  
 Qy 241 SKEIPEQRENKQKKRKVKQDKQKS 265  
 Db 241 SKEIPEQRENKQKKRKVKQDKQKS 265

RESULT 11  
 ABO44427  
 ID ABO44427 standard; protein; 265 AA.

XX ABO44427;

XX 30-SEP-2003 (first entry)

XX Human secreted protein clone DA228\_6.

XX



KW Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;  
 KW immunostimulant; vulnary; haematopoietic stem cell; gene therapy;  
 KW supporting factor for proliferation of stem cells; wound healing;  
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;  
 KW bone marrow transplantation; cord blood transplantation;  
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;  
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;  
 KW congenital anaemia; sickle cell; Gaucher's disease; morphogenesis;  
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;  
 KW cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.

OS US2003044792-A1.

PN 06-MAR-2003.

XX 28-JUN-2001; 2001US-00894912.

XX 28-JUN-2000; 2000US-0215733P.

PR 03-FEB-2001; 2001US-0286614P.

PR 05-APR-2001; 2001US-0282397P.

XX (TANG/) TANG Y T.

PA (LABA/) LABAT I.

PA (DRMA/) DRMANAC R T.

PA (MIZE/) MIZE N.

PA (NISH/) NISHIKAWA M.

PA (CHAO/) CHAO C.

XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

PI WPI; 2003-625403/59.

XX Novel isolated polypeptide having stem cell growth factor activity,  
 PT useful for promoting wound healing, and as a medicine to proliferate or  
 PT support human hematopoietic stem cells or human hematopoietic progenitor  
 PT cells.

XX Disclosure; Fig 3; 96pp; English.

XX The invention relates to an isolated stem cell growth factor-like  
 CC polypeptide (referred as supporting factor for proliferation of stem  
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or  
 CC fragment, analogue, variant or derivative, that retains stem cell growth  
 CC factor activity. Also included are an isolated polynucleotide encoding  
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or  
 CC derivative, that retains stem cell growth factor activity, or the  
 CC complement of the polynucleotide), an (expression) vector comprising the  
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the  
 CC SCR-1 polynucleotide in operative association with a regulatory sequence  
 CC that controls expression of the polynucleotide in the host cell,  
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an  
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an  
 CC activity to support proliferation or survival of haematopoietic stem cell  
 CC or haematopoietic progenitor cell, with a proviso that C-terminal aa  
 CC sequence does not comprise the aa sequence appearing as ABO44433), an  
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and  
 CC lacking any 10 consecutive aas from ABO44430, an isolated polypeptide  
 CC with stem cell growth factor activity having at least an aa sequence  
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1  
 CC polypeptide to maintain survival of or promote proliferation of a stem  
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array  
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1  
 CC polynucleotide attached to a surface, a stromal cell genetically  
 CC engineered to express the SCR-1 polypeptide to support proliferation or  
 CC survival of a stem cell or germ cell and an implant comprising a cell  
 CC genetically engineered to express the SCR-1 polypeptide to support  
 CC proliferation or survival of a stem cell or germ cell. The SCR-1  
 CC polypeptide is useful for identifying a compound that binds to the SCR-1  
 CC polypeptide and for maintaining survival of or promoting proliferation of  
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic  
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound

CC healing. The human haematopoietic stem cell or human haematopoietic  
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a  
 CC graft for the conventional bone marrow transplantation or cord blood  
 CC transplantation. The transplantation of haematopoietic stem cells can be  
 CC employed as a therapy for treating diseases such as chronic granulomatous  
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,  
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia  
 CC due to enzyme defect, congenital anaemia such as sickle cell, Gaucher's  
 CC disease etc. The SCR-1 polypeptide is useful for cell growth and  
 CC morphogenesis, including tissue specific stem cell growth, epithelial  
 CC cell growth, and regulation, ovarian follicle development, promoting nerve  
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone  
 CC growth and immunosuppression. The present sequence is a protein  
 CC homologous to Human SCR-1 protein

XX Sequence 265 AA;

Query Match 97.8%; Score 1477; DB 6; Length 265;

Best Local Similarity 100.0%; Pred. No. 3.8e-108;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISLWLFILNFMFYIGSQNASRGRQRMRHPNVSGQGGCATSDYNGCLSKR 60

Db 1 MHLRLISLWLFILNFMFYIGSQNASRGRQRMRHPNVSGQGGCATSDYNGCLSKR 60

Qy 61 LPPALERIGMKOIGVCLSSCPSSGYGTYRDPINCKTKCADCTCFNKNFCTKSGFYL 120

Db 61 LPPALERIGMKOIGVCLSSCPSSGYGTYRDPINCKTKCADCTCFNKNFCTKSGFYL 120

Qy 121 HLGLCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180

Db 121 HLGLCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180

Qy 181 QHPSAKGNLCPTTNETRKTCTVQRKKCGKRGKGRKRKPKPNKGSKEALPDSEKLS 240

Db 181 QHPSAKGNLCPTTNETRKTCTVQRKKCGKRGKGRKRKPKPNKGSKEALPDSEKLS 240

Qy 241 SKELPQRENKQKKRKVKQDKQS 265

Db 241 SKELPQRENKQKKRKVKQDKQS 265

RESULT 12

AAW85607

ID AAW85607 standard; protein; 292 AA.

AC AAW85607;

DT 02-MAR-1999 (first entry)

DE Secreted protein clone da228\_6.

XX Clone; secreted protein; protein factor; cytokine; lymphokine;  
 XX interferon; colony stimulating factor; CSF; interleukin; cloning;  
 XX tumour invasion; tumour suppression; immune boosting.

OS Homo sapiens.

PN WO9849302-A1.

PD 05-NOV-1998.

XX 24-APR-1998; 98WO-US008336.

XX 25-APR-1997; 97US-00845296.

PR 23-APR-1998; 98US-00065125.

PA (GENY ) GENETICS INST INC.

PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;  
 PI Spaulding V, Agostino MJ;

DR WPI; 1999-024059/02.



DR N-PSDB; AAV831133.

XX New polynucleotides encoding secreted human proteins - are derived from

PT human foetal brain, adult brain, adult blood or placenta cDNA libraries,

PT useful, e.g. as potential immunomodulators.

XX Claim 8; Page 63-64; 104pp; English.

XX The nucleotide sequence (NS) of the full-length protein-coding sequence

CC of clones c1254 (AAV83132), da2286 (AAV83133), du4105 (AAV83134), eh801

CC (AAV83135), er3691 (AAV83136), fh1235 (AAV83137), fm601 (AAV83138) or

CC fr4732 (AAV83139), (all clones are deposited as ATCC 98415) and the

CC proteins they encode are predicted to have biological activities which

CC would make them suitable for treating, preventing or ameliorating medical

CC conditions in humans and animals for example, tumour suppression/invasion

CC activity, immune system boosting activity. The polynucleotides are also

CC believed to be useful for gene therapy

XX

SQ Sequence 292 AA;

Query Match 97.8%; Score 1477; DB 2; Length 292;

Best Local Similarity 100.0%; Pred. NO. 4.3e-108;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHRLISWLFILNFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

DB 1 MHRLISWLFILNFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

QY 61 LFPALERIGMKQIGVCLSSCPSGYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL 120

DB 61 LFPALERIGMKQIGVCLSSCPSGYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL 120

QY 121 HLKGLDNCPEGLEANNHTMECVSIHVCSEVSENPSPCTKKGTCGFKGTETRVREII 180

DB 121 HLKGLDNCPEGLEANNHTMECVSIHVCSEVSENPSPCTKKGTCGFKGTETRVREII 180

QY 181 QHPSAGNLCPPNTETKCTVQRKKCKQGERGKKRERKKNKGSKEAIPDSKSL 240

DB 181 QHPSAGNLCPPNTETKCTVQRKKCKQGERGKKRERKKNKGSKEAIPDSKSL 240

QY 241 SKEIPEQRENKQKKRKVDKQKS 265

DB 241 SKEIPEQRENKQKKRKVDKQKS 265

RESULT 13

ID AAE13170

XX AAE13170 standard; protein; 292 AA.

XX

AC AAE13170;

XX

DT 28-JAN-2002 (first entry)

XX

XX Human SCR-1 related protein.

XX

XX Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;

XX neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;

XX vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;

XX osteopathic; immunodeficiency syndrome; chronic granulomatous disease;

XX duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;

XX acquired immune deficiency syndrome; agammaglobulinaemia; thalasasaemia;

XX Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;

XX adrenal white matter degeneration; anaemia; neurodegenerative disease;

XX Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;

XX severe combined immunodeficiency; immune disorder; autoimmune disorder;

XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;

XX autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;

XX supporting factor for the proliferation of stem cell.

XX

OS Unidentified.

XX

XX WO200177169-A2.

XX

PD 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX

XX 05-APR-2000; 2000US-00543774.

PR 28-JUN-2000; 2000US-0215733P.

PR 09-JAN-2001; 2001US-00757562.

PR 05-FEB-2001; 2001US-0266614P.

XX (HYSE-) HYSEQ INC.

PA (KIRI ) KIRIN BEER KK.

XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;

PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

XX WPI; 2001-657166/75.

DR N-PSDB; AAD21740.

XX

XX Novel stem cell growth factor like polypeptides and polynucleotides for

PT identifying modulators useful for treating diseases such as Alzheimer's

PT disease, cancer, rheumatoid arthritis, osteoporosis.

XX

PS Claim 27; Page 231-232; 232pp; English.

XX

CC The patent discloses novel stem cell growth factor-like proteins and

CC polynucleotides encoding them. Proteins of the invention are also known

CC as supporting factor for the proliferation of stem cells (SCR-1). Stem

CC cell growth factor-like proteins are useful for supporting proliferation

CC or survival of a stem cell or germ cell which is preferably primordial

CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem

CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent

CC cell. The haematopoietic progenitor cell cultured using stem cell growth

CC factor-like proteins can replace as a graft for the bone marrow

CC transplantation or cord blood transplantation for treating a variety of

CC diseases such as immunodeficiency syndrome, chronic granulomatous

CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,

CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),

CC thalasasaemia, haemolytic anaemia due to enzyme defect, congenital anaemia

CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage

CC diseases such as mucopolysaccharidosis, adrenal white matter

CC degeneration, a variety of cancer and tumours. Proteins of the invention

CC are useful for treating diseases such as Parkinson's disease, Alzheimer's

CC disease and other neurodegenerative diseases, thrombocytopaenia, immune

CC deficiencies and disorders such as severe combined immunodeficiency

CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic

CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary

CC inflammation. Sequences of the invention are also useful in gene therapy.

CC The present sequence is human SCR-1 related protein

XX

SQ Sequence 292 AA;

Query Match 97.8%; Score 1477; DB 4; Length 292;

Best Local Similarity 100.0%; Pred. NO. 4.3e-108;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHRLISWLFILNFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

DB 1 MHRLISWLFILNFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

QY 61 LFPALERIGMKQIGVCLSSCPSGYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL 120

DB 61 LFPALERIGMKQIGVCLSSCPSGYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL 120

QY 121 HLKGLDNCPEGLEANNHTMECVSIHVCSEVSENPSPCTKKGTCGFKGTETRVREII 180

DB 121 HLKGLDNCPEGLEANNHTMECVSIHVCSEVSENPSPCTKKGTCGFKGTETRVREII 180

QY 181 QHPSAGNLCPPNTETKCTVQRKKCKQGERGKKRERKKNKGSKEAIPDSKSL 240

DB 181 QHPSAGNLCPPNTETKCTVQRKKCKQGERGKKRERKKNKGSKEAIPDSKSL 240

QY 241 SKEIPEQRENKQKKRKVDKQKS 265

Db 241 SKEIPEQRENKQKKRKVKQDKQS 265

RESULT 14  
ABP61846  
ID ABP61846 standard; protein; 292 AA.  
AC ABP61846;  
XX  
DT 04-OCT-2002 (first entry)  
XX  
DB Human polypeptide SEQ ID NO 200.  
XX  
KW Human; cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;  
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;  
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;  
KW anticancer; fungicide; antidiabetic; antilashmatic; antiallergic;  
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;  
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;  
KW stem cell; growth factor; nervous system disease; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW osteoporosis; severe combined immunodeficiency; SCID; infection;  
KW multiple sclerosis; rheumatoid arthritis; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FN US2002065394-A1.  
XX  
PD 30-MAY-2002.  
XX  
PF 22-DEC-2000; 2000US-00745763.  
XX  
PR 18-MAR-1998; 98US-00040963.  
XX  
PA (JACO/) JACOBS K.  
PA (MCCO/) MCCOY J M.  
PA (LAVA/) LAVALLIE E R.  
PA (COLL/) COLLINS-RACIE L A.  
PA (EVAN/) EVANS C.  
PA (MERB/) MERBERG D.  
PA (TREA/) TREACY M.  
PA (SPAU/) SPAULDING V.  
XX  
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Spaulding V;  
XX  
DR WPI: 2002-582343/62.  
DR N-PSDB; ABQ92060.  
XX  
PT Novel secreted or transmembrane protein and polynucleotide encoding the  
PT protein, useful for diagnosis and treatment of neurological disorders,  
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.  
XX  
PS Claim 207; Page 203-204; 284pp; English.  
XX  
CC The invention relates to human secreted or transmembrane protein (I),  
CC their fragments and is encoded by specific complementary deoxyribonucleic  
CC acid (cDNA) inserts (II), where the protein is substantially free from  
CC other mammalian proteins. (I) are useful for preventing, treating or  
CC ameliorating a medical condition, especially immunological treatment or  
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,  
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,  
CC stem cell growth factor activity and activin or inhibin-related  
CC activities. (I) can be used to manipulate stem cells in culture to give  
CC rise to neuroepithelial cells that can be used to augment or replace  
CC cells damaged by illness, autoimmune disease, accidental damage or  
CC genetic disorders. (I) induces the proliferation of neural cells and  
CC regeneration of nerve and brain tissue and is useful for the treatment of  
CC central and peripheral nervous system diseases and neuropathies, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis and is useful for treating myeloid  
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia

CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,  
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or  
CC periodontal disease. (I) is also useful for gut protection or  
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury  
CC in various tissues, various immune deficiencies and disorders including  
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,  
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,  
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,  
CC such as asthma or other respiratory problems. (II) is useful to express  
CC recombinant protein, as markers for tissues in which the corresponding  
CC protein is preferentially expressed and in gene therapy. The present  
CC sequence is that of a polypeptide of the invention  
XX

SQ Sequence 292 AA;

Query Match 97.8%; Score 1477; DB 5; Length 292;  
Best Local Similarity 100.0%; Pred. No. 4.3e-108;  
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHLRLISWLFILNPMYIGSONASGRGRORRMHPNVSGCGGCATCSDYNGCLSKPR 60  
DB 1 MHLRLISWLFILNPMYIGSONASGRGRORRMHPNVSGCGGCATCSDYNGCLSKPR 60  
QY 61 LFFALERIGMKQIGVCLSSCPGSGYGTVPDINKCTKCKADCDTCFNKNFCTCKSGFYL 120  
DB 61 LFFALERIGMKQIGVCLSSCPGSGYGTVPDINKCTKCKADCDTCFNKNFCTCKSGFYL 120  
QY 121 HLGKCLDNCPEGLBANNHTMECVSIHVCEVSEWNPMSPTCKGKTCGPKRGTTREII 180  
DB 121 HLGKCLDNCPEGLBANNHTMECVSIHVCEVSEWNPMSPTCKGKTCGPKRGTTREII 180  
QY 181 QHPSAKGNLCPTTNETRKTCTVQRKKCKQGERGKGRERKPKPNKGSKEAIPDSKLS 240  
DB 181 QHPSAKGNLCPTTNETRKTCTVQRKKCKQGERGKGRERKPKPNKGSKEAIPDSKLS 240  
QY 241 SKEIPEQRENKQKKRKVKQDKQS 265  
DB 241 SKEIPEQRENKQKKRKVKQDKQS 265

RESULT 15

ABR62114  
ID ABR62114 standard; protein; 292 AA.  
XX  
AC ABR62114;  
XX  
DT 18-AUG-2003 (first entry)  
XX  
DE Human secreted protein clone da\_288\_6 #25.

XX Human; secreted; stem cell growth factor; cytostatic; haemostatic;  
KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;  
KW degenerative disease; Alzheimer's disease; food supplement;  
KW immunological disorder.

XX Homo sapiens.

XX WO2003029405-A2.

XX 10-APR-2003.

XX 30-AUG-2002; 2002WO-US027746.

XX 30-AUG-2001; 2001US-0316368P.

PR 10-DEC-2001; 2001US-0339739P.

PR 19-APR-2002; 2002US-00125852.

XX (HYSE-) HYSEQ INC.

XX Tang YT;

DR WPI; 2003-381616/36.

XX New stem cell growth factor-like polypeptides and polynucleotides, useful  
PT for treating e.g. leukemia, hemophilia and degenerative diseases like  
PT Alzheimer's disease, and for inducing immune response.  
XX  
PS Disclosure; Fig 2; 151pp; English.  
PS  
XX The invention relates to new stem cell growth factor-like polypeptides  
CC and polynucleotides. The stem cell growth factor-like polypeptides and  
CC polynucleotides are useful for inducing differentiation of embryonic and  
CC adult stem cells to give rise to different cell types, for treating e.g.  
CC leukaemia, haemophilia and degenerative diseases like Alzheimer's  
CC disease. They are also useful for generating new tissues and organs that  
CC may aid patients in need of transplanted tissues. The polynucleotides are  
CC useful as hybridisation probes, oligomers or primers for PCR, for  
CC chromosome and gene mapping, in recombinantly producing protein, in  
CC generating antisense DNA or RNA, in diagnostics as expressed sequence  
CC tags for identifying expressed genes, and for inducing immune response.  
CC The polypeptides are useful for generating antibodies that specifically  
CC bind the polypeptide, as molecular weight markers, and as a food  
CC supplement (e.g. protein or amino acid supplement, and as a carbon,  
CC nitrogen or carbohydrate source). Compositions comprising the  
CC polypeptides or polynucleotides are useful for the diagnosis, treatment  
CC or prevention of cancers, and other immunological disorders. The current  
CC sequence represents a human secreted protein clone da\_288\_6  
XX  
SQ Sequence 292 AA;  
Query Match 97.8%; Score 1477; DB 6; Length 292;  
Best Local Similarity 100.0%; Pred. No. 4.3e-108;  
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHLRLISWLPFIILNFMETIGSQNASRGRORRHPNVSGCGGCATCSDYNGCLSCKPR 60  
DB 1 MHLRLISWLPFIILNFMETIGSQNASRGRORRHPNVSGCGGCATCSDYNGCLSCKPR 60  
QY 61 LFFALERIGMKOIGVCLSSCPGSGYCTRYPDINKCTKACDCTCFNKPFCTKCKSGPYL 120  
DB 61 LFFALERIGMKOIGVCLSSCPGSGYCTRYPDINKCTKACDCTCFNKPFCTKCKSGPYL 120  
QY 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPMSPTCKGKTCGFKRGTTETRVREII 180  
DB 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPMSPTCKGKTCGFKRGTTETRVREII 180  
QY 181 QHPSAKGNLCPTNTRKCTVQRKKQKGERGKGRERKRKPKNKGESKEAIPDSKSLES 240  
DB 181 QHPSAKGNLCPTNTRKCTVQRKKQKGERGKGRERKRKPKNKGESKEAIPDSKSLES 240  
QY 241 SKEIPEORENKQOKKRVQDKQS 265  
DB 241 SKEIPEORENKQOKKRVQDKQS 265

Search completed: June 29, 2004, 16:59:53  
Job time : 50.5178 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:57:05 ; Search time 14.1942 Seconds  
(without alignment)  
989.298 Million cell updates/sec

Title: US-09-894-912A-34

Perfect score: 1510

Sequence: 1 MHLRLISWLPILNPFMYIG.....QQKKRKVDKQKSVSVTVH 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	173.5	11.5	969	2	US-08-284-941-2
2	173.5	11.5	969	2	US-08-447-642-2
3	173.5	11.5	969	3	US-09-236-503-2
4	173.5	11.5	969	5	PCT-US93-02147A-2
5	162	10.7	799	2	US-08-525-940-23
6	162	10.7	799	2	US-08-976-838-23
7	162	10.7	881	2	US-08-525-940-21
8	162	10.7	881	2	US-08-976-838-21
9	162	10.7	915	2	US-08-525-940-18
10	162	10.7	915	2	US-08-976-838-18
11	162	10.7	915	4	US-09-214-555B-7
12	162	10.7	915	4	US-09-214-555B-2
13	161	10.7	288	1	US-08-368-852-15
14	158.5	10.5	379	4	US-09-907-794A-4
15	158.5	10.5	379	4	US-09-905-125A-4
16	158.5	10.5	379	4	US-09-902-775A-4
17	157.5	10.4	288	2	US-08-525-940-15
18	157.5	10.4	288	2	US-08-976-838-15
19	150.5	10.0	380	4	US-09-205-258-441
20	149	9.9	568	1	US-07-862-021B-14
21	149	9.9	568	5	PCT-US93-03164-14
22	148.5	9.8	2523	1	US-08-185-432-18
23	148.5	9.8	2523	4	US-08-899-232-3
24	147.5	9.8	802	1	US-07-862-021B-12
25	147.5	9.8	802	1	US-08-313-288B-12
26	147.5	9.8	802	5	PCT-US93-03164-12
27	146.5	9.7	807	4	US-09-132-769-1

28	146.5	9.7	807	4	US-09-640-173-186	Sequence 186, App
29	146.5	9.7	807	4	US-09-713-550-186	Sequence 186, App
30	146	9.7	1068	1	US-08-537-210A-2	Sequence 2, Appli
31	146	9.7	1068	3	US-09-113-825-2	Sequence 2, Appli
32	146	9.7	2556	1	US-08-185-432-17	Sequence 17, Appl
33	146	9.7	2556	4	US-08-899-232-2	Sequence 2, Appli
34	144	9.5	807	4	US-09-132-769-3	Sequence 3, Appli
35	143.5	9.5	807	1	US-07-862-021B-10	Sequence 10, Appl
36	143.5	9.5	807	1	US-08-313-288B-10	Sequence 10, Appl
37	143.5	9.5	807	4	US-09-132-769-5	Sequence 5, Appli
38	143.5	9.5	807	5	PCT-US93-03164-10	Sequence 103, App
39	139	9.2	366	3	US-08-857-076-103	Sequence 5, Appli
40	139	9.2	486	3	US-08-746-559A-5	Sequence 4, Appli
41	139	9.2	516	3	US-08-746-559A-4	Sequence 2, Appli
42	1367	9.2	1367	2	US-08-249-687C-2	Sequence 2, Appli
43	139	9.2	1367	2	US-08-625-819-2	Sequence 2, Appli
44	139	9.2	1367	3	US-08-746-559A-2	Sequence 2, Appli
45	139	9.2	1367	4	US-08-864-641B-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-08-284-941-2  
; Sequence 2, Application US/08284941  
; Patent No. 5863756  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; APPLICANT: KIEFER, MICHAEL C  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACR 4 AND  
; TITLE OF INVENTION: PACR 4.1 GENE AND POLYPEPTIDES IN CELLS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/284,941  
; FILING DATE: 2 August 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEELY PH.D., RICHARD L.  
; REGISTRATION NUMBER: 30092  
; REFERENCE/DOCKET NUMBER: CHIR-009/01US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 843-5070  
; TELEFAX: (415) 857-0663  
; TELEX: 380816 COOLEY PA  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 969 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULAR TYPE: protein  
US-08-284-941-2

Query Match 11.5%; Score 173.5; DB 2; Length 969;  
Best Local Similarity 28.8%; Pred. No. 1.3e-06;  
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;  
Oy 18 YIGSQNASRGRRMRHNPVNSQGGCGCATCS--YNGCLCKPLPFALEIRGMKQIGV 75  
Db 737 YFGDTAARRCR-----CHKGCETCSRATQCLSCR-RGFY-----HHQEMNT 779

QY 76 CLSCPSGYGTRYPDINKCTCKKADCDTCFNK-NFCTCKSGFYHLGKLDNCPGLE 134  
Db 780 CVTLCPAGFYADB--SQNCLKHPSCCKVDEPEKCTVCKBGFSLARGSCIPDCBPGTY 837  
QY 135 ANNHTECVSIVHCEVSEWNPSPCTKKGKTG 167  
Db 838 FDSELRGEGCH-----TCG 853

RESULT 2  
US-08-447-642-2  
; Sequence 2, Application US/08447642  
; Patent No. 5989890  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND  
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/447,642  
; FILING DATE: 23-MAY-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/284,941  
; FILING DATE: 2 August 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEELEY PH.D., RICHARD L.  
; REGISTRATION NUMBER: 30092  
; REFERENCE/DOCKET NUMBER: CHIR-009/0105  
; TELEPHONE: (415) 843-5070  
; TELEFAX: (415) 857-0663  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 969 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-447-642-2

Query Match 11.5%; Score 173.5; DB 2; Length 969;  
Best Local Similarity 28.8%; Pred. No. 1.3e-06;  
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 18 YIGSQNASRRQRMRHPNVSGCGGCATCSD--YNGCLSKPRLPFALERIGMKQIGV 75  
Db 737 YFGDTAARRCRR-----CHKGCTCSSRAATQCLSCR-RGFFY-----HHQEMNT 779

QY 76 CLSCPSGYGTRYPDINKCTCKKADCDTCFNK-NFCTCKSGFYHLGKLDNCPGLE 134  
Db 780 CVTLCPAGFYADB--SQNCLKHPSCCKVDEPEKCTVCKBGFSLARGSCIPDCBPGTY 837

QY 135 ANNHTECVSIVHCEVSEWNPSPCTKKGKTG 167  
Db 838 FDSELRGEGCH-----TCG 853

; Sequence 2, Application US/09236503  
; Patent No. 6277590  
; GENERAL INFORMATION:  
; APPLICANT: Barr, Philip J  
; TITLE OF INVENTION: Compositions and Methods for PACE 4 and 4.1 Gene and  
; TITLE OF INVENTION: Polypeptides in Cells  
; FILE REFERENCE: CHIR-009/04US  
; CURRENT APPLICATION NUMBER: US/09/236,503  
; CURRENT FILING DATE: 1999-01-25  
; EARLIER APPLICATION NUMBER: 08/447,642  
; EARLIER FILING DATE: 1995-05-23  
; EARLIER APPLICATION NUMBER: 08/284,941  
; EARLIER FILING DATE: 1994-08-02  
; EARLIER APPLICATION NUMBER: 07/848,629  
; EARLIER FILING DATE: 1992-03-09  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 969  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-236-503-2

Query Match 11.5%; Score 173.5; DB 3; Length 969;  
Best Local Similarity 28.8%; Pred. No. 1.3e-06;  
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 18 YIGSQNASRRQRMRHPNVSGCGGCATCSD--YNGCLSKPRLPFALERIGMKQIGV 75  
Db 737 YFGDTAARRCRR-----CHKGCTCSSRAATQCLSCR-RGFFY-----HHQEMNT 779

QY 76 CLSCPSGYGTRYPDINKCTCKKADCDTCFNK-NFCTCKSGFYHLGKLDNCPGLE 134  
Db 780 CVTLCPAGFYADB--SQNCLKHPSCCKVDEPEKCTVCKBGFSLARGSCIPDCBPGTY 837

QY 135 ANNHTECVSIVHCEVSEWNPSPCTKKGKTG 167  
Db 838 FDSELRGEGCH-----TCG 853

RESULT 4  
PCT-US93-02147A-2  
; Sequence 2, Application PC/TUS9302147A  
; GENERAL INFORMATION:  
; APPLICANT: BARR, PHILIP J  
; APPLICANT: KIEFER, MICHAEL C  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND  
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/02147A  
; FILING DATE: 19930309  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/848,629  
; FILING DATE: 09-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEELEY PH.D., RICHARD L  
; REGISTRATION NUMBER: 30092  
; REFERENCE/DOCKET NUMBER: CHIR-009/0005

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLRY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-02147A-2

Query Match 11.5%; Score 173.5; DB 5; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.3e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 18 YIGSNASRGRRQRMHPNVSGCGGATCSD--YNGCLCKPLPFALEIRGMKQIGV 75
Db 737 YFGDTAARRCR-----CHKGCETCSRAATQCLSCR-RGFY-----HHQEMNT 779

QY 76 CLSSCPGYYGTRYPDINKCTCKADCTCFNK-NFTYCKSGFYHLGKCLDNCPEGLE 134
Db 780 CVTLCPAGFYADE--SQKNCLKPSCKKCVDEPEKCTVCKGFGSLARGSCIPDCPEGT 837

QY 135 ANNHTECVSIHVCEVSEWNPSPCTKKGKTCG 167
Db 838 FDSILRCBCH-----TCG 853

RESULT 5
US-08-525-940-23
; Sequence 23, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLRY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-02147A-2

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.8e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY 20 GSONASRGRRQRMHPNVSGCGGATCSDYNG--CLSKPLPFALEIRGMKQIGVCL 77
Db 563 GHYHADK-KRCRKCAPN-----CESCFGSHGDDQNSCKGYFL-----NEETNSCV 607

QY 78 SSCPSG-YGTRYPDINKCTCKADCTCFNK-NFTYCKSGFYHLGKCLDNCPEGLEAN 136
Db 608 THCPDGSYQDTYK--NLCKKSENCKYCTEHPNCTECDGLSLOGSRCSVSCDGYFN 664

QY 137 NHTMB-----CVS-----IVHCEVSEW-----NPWSPCTKK 162
Db 665 GQDCQPCHRFCATCAGAGAGCGINCTEGYFMEDEGRCVQSCSISYFDSHSENGYKSKKC 724

QY 163 GHTC-----GKRGTE-----TRVPELHQPSAKGNLCPTTNETRKTCTV 201
Db 725 DISLTCNGPGPKNCTCSPSGYLLDGMCMGAICKDAKTESSWAEGGFCMLVKNNILC-- 782

QY 202 QRKKCQK 208
Db 783 QRKVLQQ 789

RESULT 6
US-08-976-838-23
; Sequence 23, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.8e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
```





[illegible]

```
QY 202 QKCKCK 208
Db 899 QKVLQO 905

Query Match 10.7%; Score 162; DB 4; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

RESULT 11
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match 10.7%; Score 162; DB 4; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY 20 GSONASRGRRQRHMHNVSGCGGCGATCSDYNG--CLSCCKPLFFALERIGMKQIGVCL 77
Db 679 GHYHADK-KRCRKAPE-----CESCFGSHGDCMCKYGVFL-----NEETNSCV 723

QY 78 SSCPSG-YGYTRYPDINKTKKADCDYCFNKNFCTKCKSGFYHLGKCLDNCPEGLEAN 136
Db 724 THCPDSGYQDTTK--NLCKKCSNCKYCTEFNCTECDGLSLQGRSCSVSCEDGRYFN 780

QY 137 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 162
Db 781 GDCQPCCHRPFATCAGAGADGCTEGYFMEDGRCVQSCSIYYPHSHSENGYKSKCKC 840

QY 163 GKTC-----GPKRGTE-----TRVREIIQHPSAKGNLCPTTNETRKTCTV 201
Db 841 DISCLTCNGPGFNCTSCPSGYLLDLGMQMGAIKDAEESWABGGFCMLVKKNLNC-- 898

QY 202 QKCKCK 208
Db 899 QKVLQO 905

Query Match 10.7%; Score 161; DB 1; Length 288;
Best Local Similarity 28.2%; Pred. No. 3.8e-06;
Matches 46; Conservative 25; Mismatches 52; Indels 40; Gaps 11;

QY 35 PNVSQ-GCGG-GCATCSDYNGCLSCKPLFFALERIGMKQIGVCLSCSPSGYGYTRYPD 92

Query Match 10.7%; Score 162; DB 4; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

RESULT 12
US-09-214-555B-7
; Sequence 7, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-7
```

Db 13 PECSVGCDCGPDHCND---CL-----HYVYKLG-----NNTRICVSSCPGHHY---HADK 58  
Qy 93 NKCTKACADCTCFNK--NFTCKKSGFYHLH--LGKCLDNCPEGLEANNHTMBCVSIYHC 148  
Db 59 KRCRCAPNCSCFSGHGDQCMCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 112  
Qy 149 EVSEWNPMSPTCKGKTC-GFKRGTTETRVREIIQHPSAKGNLC 190  
Db 113 -----RKSENKKUTETFAKTECR-----DGLSLQGSRC 142

## RESULT 14

US-09-907-794A-4

; Sequence 4, Application US/09907794A  
; Patent No. 663468  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas P.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/907,794A  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: PCT/US90/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095

; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 4  
; LENGTH: 379  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-907-794A-4

Query Match 10.5%; Score 158.5; DB 4; Length 379;

Best Local Similarity 24.9%; Pred. No. 8.5e-06;  
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Qy 41 CQGCAT---CSDVNGC-----LSCKRLPFALERIGMKQIGVCLSS---C 80  
Db 182 CPGCGNGGPNERRICEPCDPGPHCEKALCTPCMG-----GLCVTPGFCIC 232  
Qy 81 PSYGYTRYDPINKTKC-KADCD-TCFNKNFTCKKSGFYHLGKCLDNCPEGLEANNH 138  
Db 233 PPGFYG-----VNCCKANCSFTCFNGGTC-----FY--PGKCI--CPPGLEGE-- 271  
Qy 139 TMECVSIVHCEVSEWNPMSPTCKGKTCGFKRGTTETRVREIIQHPSAKGNLCPP----- 192  
Db 272 -----QCHSKCP--QPCRNGGKCTG--KSKCKSGKYQGLCKSPVCEPGCGAHG 318  
Qy 193 -TNETRKTQVQKKCKGKRGKGRKRRK-----KPNKGSKEAIPDSKLSLESSKEIP 245  
Db 319 TCHPKNK-----CQCEGWHGRHCNKRYEASLIHALRPAQALRQHTTFLSKAEERDDPP 373  
Qy 246 E 246  
Db 374 E 374

## RESULT 15

US-09-905-125A-4

; Sequence 4, Application US/09905125A  
; Patent No. 6664376  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas P.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/905,125A

```
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 4
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-905-125A-4

Query Match      10.5%; Score 158.5; DB 4; Length 379;
Best Local Similarity 24.9%; Pred. No. 8.5e-06;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

QY      41  CQGCAT---CSDYNGC-----LSCKPLFFALERIGMKQIGVCLSS-----C 80
Db      182  CPGGCRNGGFCNERRICEPCDPGPHGHCALCTPRCNG-----GLCVTPGFCIC 232

QY      81  PSGYGTRYDPINKCTKC-KADCD-TCFNKNFCTKCKSGFYHLHGKLDNCPGLEANNH 138
Db      233  PPGFYG-----VNCDKANCSCTTCFNGTC-----PY--FGKCI--CPPGLEGE-- 271

QY      139  TMECVSIHCRVSEMNPNWSCTKKGCTGPKRGTVETRVREIIQHPSAKGNLCP----- 192
Db      272  -----QCISKCP--QPCRNKGKCIQ---KSKCKSGYQDGLCKSPVCFPGCGGANG 318

QY      193  -TNETRKTQVKKCKGKGRKRRK-----KFNKGESKKAIPDSKSLSSKHIP 245
Db      319  TCHFPKNC-----CQSGWHGRHCNRYEASLIHALRPAQAQLRQHTPSLAKAERRDPP 373

QY      246  E 246
Db      374  E 374
```

Search completed: June 29, 2004, 17:04:02  
Job time : 15.1942 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 17:02:26 ; Search time 36.6408 Seconds  
(without alignments)  
2098.641 Million cell updates/sec

Title: US-09-894-912A-34  
Perfect score: 1510  
Sequence: 1 MHLRLISWLFILNFMFYIG.....QQKKRKVDKQKSVSVTVH 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubpa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1510	100.0	272	10	US-09-894-912A-10
2	1510	100.0	272	10	US-09-894-912A-34
3	1510	100.0	272	14	US-10-125-852-23
4	1505	99.7	273	10	US-09-894-912A-13
5	1477	97.8	265	10	US-09-894-912A-26
6	1477	97.8	292	9	US-09-745-763-166
7	1477	97.8	292	10	US-09-894-912A-48
8	1477	97.8	292	14	US-10-125-852-25
9	1400	92.7	251	10	US-09-894-912A-16
10	1315.5	87.1	279	10	US-09-894-912A-32
11	1280	84.8	225	14	US-10-185-770-4
12	1218	80.7	239	12	US-10-087-192-1782
13	1183	78.3	239	15	US-10-094-886-172
14	1111	73.6	239	15	US-10-094-886-176
15	1107	73.3	195	15	US-10-094-886-174

16	981	65.0	180	12	US-10-087-192-1779	Sequence 1779, Ap
17	903	59.8	160	10	US-09-894-912A-14	Sequence 14, Appl
18	903	59.8	160	12	US-10-276-774-1744	Sequence 1744, Ap
19	656	43.4	263	14	US-10-125-852-18	Sequence 18, Appl
20	648.5	42.9	243	14	US-10-125-852-21	Sequence 21, Appl
21	644	42.6	265	14	US-10-125-852-24	Sequence 24, Appl
22	638	42.3	229	10	US-09-894-912A-25	Sequence 25, Appl
23	626.5	41.5	236	15	US-10-108-260A-4829	Sequence 4829, Ap
24	589.5	39.0	243	14	US-10-185-770-2	Sequence 2, Appli
25	582.5	38.6	243	14	US-10-125-852-13	Sequence 13, Appl
26	582.5	38.6	243	16	US-10-467-042-12	Sequence 12, Appl
27	578	38.3	250	14	US-10-125-852-3	Sequence 3, Appli
28	545	36.1	222	14	US-10-125-852-15	Sequence 15, Appl
29	540.5	35.8	229	14	US-10-125-852-6	Sequence 6, Appli
30	402.5	26.7	190	12	US-10-087-192-873	Sequence 873, App
31	365.5	24.2	161	12	US-10-087-192-876	Sequence 876, App
32	276	18.3	46	10	US-09-894-912A-18	Sequence 18, Appl
33	239.5	15.9	131	14	US-10-125-852-9	Sequence 9, Appli
34	221	14.6	42	10	US-09-894-912A-22	Sequence 22, Appl
35	213	14.1	37	10	US-09-894-912A-20	Sequence 20, Appl
36	202	13.4	110	14	US-10-125-852-11	Sequence 11, Appl
37	186.5	12.4	1548	14	US-10-180-903-2	Sequence 2, Appli
38	177.5	11.8	43	14	US-10-125-852-7	Sequence 7, Appli
39	173.5	11.5	969	10	US-09-961-403-6	Sequence 6, Appli
40	160.5	10.6	337	12	US-10-357-820-20	Sequence 20, Appl
41	160.5	10.6	343	12	US-10-357-820-8	Sequence 8, Appli
42	160.5	10.6	365	12	US-10-357-820-4	Sequence 4, Appli
43	160.5	10.6	365	12	US-10-357-820-6	Sequence 6, Appli
44	160.5	10.6	373	12	US-10-357-820-10	Sequence 10, Appl
45	160.5	10.6	479	9	US-09-764-898-221	Sequence 221, App

ALIGNMENTS

RESULT 1

US-09-894-912A-10  
; Sequence 10, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES  
; FILE REFERENCE: 28110/37260A  
; CURRENT APPLICATION NUMBER: US/09/894,912A  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/266,614  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/215,733  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 09/757,562  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 09/543,774  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-10

Query Match 100.0%; Score 1510; DB 10; Length 272;  
Best Local Similarity 100.0%; Pred. No. 5.2e-115;  
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MHLRLISWLFILNFMFYIGSNASRGRRRMRHNPVSGCGGCATCDYNGCLSCRPR	60
Db	1	MHLRLISWLFILNFMFYIGSNASRGRRRMRHNPVSGCGGCATCDYNGCLSCRPR	60
Qy	61	LPFALLERIGMKQIGVGLSSCPGYYGYTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL	120

```
Db 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Qy 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
Db 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKKRKKPKNKGSKEAIPDSKSL 240
Db 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKKRKKPKNKGSKEAIPDSKSL 240
Qy 241 SKIPEORENKQKKRKKVQDKQKSVSVTVH 272
Db 241 SKIPEORENKQKKRKKVQDKQKSVSVTVH 272
```

## RESULT 2

```
US-09-894-912A-34
; Sequence 34, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR FILING DATE: 2001-04-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2001-01-09
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-34
```

```
Query Match 100.0%; Score 1510; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 5.2e-115;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMVEYIGSNASRGRORRHPNVSQGGCGCATCSDYNGCLSCKPR 60
Db 1 MHLRLISWLFILNFMVEYIGSNASRGRORRHPNVSQGGCGCATCSDYNGCLSCKPR 60
Qy 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Qy 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
Db 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKKRKKPKNKGSKEAIPDSKSL 240
Db 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKKRKKPKNKGSKEAIPDSKSL 240
Qy 241 SKIPEORENKQKKRKKVQDKQKSVSVTVH 272
Db 241 SKIPEORENKQKKRKKVQDKQKSVSVTVH 272
```

## RESULT 3

```
US-10-125-852-23
; Sequence 23, Application US/10125852
```

```
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-
; FILE REFERENCE: HYS-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-23
```

```
Query Match 100.0%; Score 1510; DB 14; Length 272;
Best Local Similarity 100.0%; Pred. No. 5.2e-115;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMVEYIGSNASRGRORRHPNVSQGGCGCATCSDYNGCLSCKPR 60
Db 1 MHLRLISWLFILNFMVEYIGSNASRGRORRHPNVSQGGCGCATCSDYNGCLSCKPR 60
Qy 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Qy 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
Db 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKKRKKPKNKGSKEAIPDSKSL 240
Db 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKKRKKPKNKGSKEAIPDSKSL 240
Qy 241 SKIPEORENKQKKRKKVQDKQKSVSVTVH 272
Db 241 SKIPEORENKQKKRKKVQDKQKSVSVTVH 272
```

## RESULT 4

```
US-09-894-912A-13
; Sequence 13, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-13
```

Query Match	99.7%;	Score 1505;	DB 10;	Length 273;
Best Local Similarity	100.0%;	Pred. No. 1.3e-114;		
Matches 271;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	2	HLRLISLWFLFILNFMFYIGSSQASRGRQRMRMPNVSSQGGGCATCSDYNGCLSCCKPRL	61	
Db	3	HLRLISLWFLFILNFMFYIGSSQASRGRQRMRMPNVSSQGGGCATCSDYNGCLSCCKPRL	62	
Qy	62	FFALRLIGMKQIGVCLSSCSFGYGYTRYPDINCKTKCADCTCFKNKFNCTKCKSGFYHL	121	
Db	63	FFALRLIGMKQIGVCLSSCSFGYGYTRYPDINCKTKCADCTCFKNKFNCTKCKSGFYHL	122	
Qy	122	LGKCLDNCPEGLEANNHTMFCVSIHVHCEVSEWNPWSPCTKKGKTCGFKRGTEVRIIQ	181	
Db	123	LGKCLDNCPEGLEANNHTMFCVSIHVHCEVSEWNPWSPCTKKGKTCGFKRGTEVRIIQ	182	
Qy	182	HPSAKGNLCPTTNETRKTCTVORKKQKGBRGKKGRERKRKPKNKGSKEAIPDSKSLESS	241	
Db	183	HPSAKGNLCPTTNETRKTCTVORKKQKGBRGKKGRERKRKPKNKGSKEAIPDSKSLESS	242	
Qy	242	KXIPBQRNKQOQKKRKVKQDKQKSVSVSTVH	272	
Db	243	KXIPBQRNKQOQKKRKVKQDKQKSVSVSTVH	273	

```

RESULT 5
US-09-894-912A-26
; Sequence 26, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-26

```

	Query Match	97.8%	Score 1477;	DB 10;	Length 265;
	Best Local Similarity	100.0%;	Prod. No. 2.5e-112;		
	Matches 265;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MHLRLISWLFII	NFMFMEYISQNASRGRQRMRHPNVISQGGGCATCSDYNGCLCKKPR	60	
DB	1	MHLRLISWLFII	NFMFMEYISQNASRGRQRMRHPNVISQGGGCATCSDYNGCLCKKPR	60	
QY	61	LPFALERIGMKQIGVCLSSCPSGGYGYTRYPDINKCTKCADCDTCFNKNFCTCKCKSGFYLL	120		
DB	61	LPFALERIGMKQIGVCLSSCPSGGYGYTRYPDINKCTKCADCDTCFNKNFCTCKCKSGFYLL	120		
QY	121	HLKCLDNCPEGLEANNHTBECUSIVICEVSEWNPWSPCTKKGTGCGPKGTGTRVREII	180		
DB	121	HLKCLDNCPEGLEANNHTBECUSIVICEVSEWNPWSPCTKKGTGCGPKGTGTRVREII	180		
QY	181	QHP- <del>SA</del> GNLC <del>LP</del> PTN <del>ET</del> RKCTVQRKKCGKGGKGRBRKKPKNPKGSKCAIPDPSKLSL	240		

```

Db 181 OHSAGKGNLCPTNVTETKCTVQRKKCKQGBRGKKGRKKRKKKPKNKGSKKKAIPDSKSLSS 244
Qy 241 SKEIPQRENKQKKKKRKVDQKKS 265
   |||||
Db 241 SKEIPQRENKQKKKKRKVDQKKS 265
   |||||

RESULT 6
US-09-745-763-166
; Sequence 166, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; LaVallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745.763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 166:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-745-763-166

```

Query Match	97.8%	Score 1477;	DB 9;	Length 292;
Best Local Similarity	100.0%;	Prod. No. 2.7e-112;		
Matches 265;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MHLRLISLWPIIILNFMEYIGSQNASRRRRMRHPNVSGQGCGCATCSDYNGCLSKCPK 60		
Db	1	MHLRLISLWPIIILNFMEYIGSQNASRRRRMRHPNVSGQGCGCATCSDYNGCLSKCPK 60		
Qy	61	LFPALEIRIGMKQIGVCLSSCPSGYGYTRYDPDIINKCTKACADCDTCFKNKFTCKSGFYL 120		
Db	61	LFPALEIRIGMKQIGVCLSSCPSGYGYTRYDPDIINKCTKACADCDTCFKNKFTCKSGFYL 120		
Qy	121	HLGKCLDNCPEGLAANNHTMECVSIHVCEYSEWNPSPCTCKGKTCQFKGTEVRVEII 180		
Db	121	HLGKCLDNCPEGLAANNHTMECVSIHVCEYSEWNPSPCTCKGKTCQFKGTEVRVEII 180		
Qy	181	QHPISAKGNLCPPNETHKCTVQRKKCKGRRGKGRKRRKQPNKGSKEAI PDSKELES 240		



Db 181 QHPKAGNLCPTNTRKCTVQRKCKGGRGKGRKRRKPKNKGSKEAIPDSKSL 240  
Qy 241 SKEIPEQRENKQKKRKKVQDKOKS 265  
Db 241 SKEIPEQRENKQKKRKKVQDKOKS 265

## RESULT 7

US-09-894-912A-48  
; Sequence 48, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 48  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-48

Query Match 97.8%; Score 1477; DB 10; Length 292;  
Best Local Similarity 100.0%; Pred. No. 2.7e-112; Indels 0; Gaps 0;  
Matches 265; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0;  
Qy 1 MHLRLISWLFILNFMPIYIGSNASRGRQRMRHMPNVQSCGGCATCSDYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMPIYIGSNASRGRQRMRHMPNVQSCGGCATCSDYNGCLSKPR 60  
Qy 61 LPPALERIGMKQIGVCLSSCPGTYGTRYPDINKCTKCKADCDTCFKNKFCCKSGPYL 120  
Db 61 LPPALERIGMKQIGVCLSSCPGTYGTRYPDINKCTKCKADCDTCFKNKFCCKSGPYL 120  
Qy 121 HLGKCLDNCPEGLNANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180  
Db 121 HLGKCLDNCPEGLNANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180  
Qy 181 QHPKAGNLCPTNTRKCTVQRKCKGGRGKGRKRRKPKNKGSKEAIPDSKSL 240  
Db 181 QHPKAGNLCPTNTRKCTVQRKCKGGRGKGRKRRKPKNKGSKEAIPDSKSL 240  
Qy 241 SKEIPEQRENKQKKRKKVQDKOKS 265  
Db 241 SKEIPEQRENKQKKRKKVQDKOKS 265

## RESULT 8

US-10-125-852-25  
; Sequence 25, Application US/10125852  
; Publication No. US20030032034A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES  
; FILE REFERENCE: HYS-43A  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: US/10/125,852  
; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-125-852-25

Query Match 97.8%; Score 1477; DB 14; Length 292;  
Best Local Similarity 100.0%; Pred. No. 2.7e-112; Indels 0; Gaps 0;  
Matches 265; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0;  
Qy 1 MHLRLISWLFILNFMPIYIGSNASRGRQRMRHMPNVQSCGGCATCSDYNGCLSKPR 60  
Db 1 MHLRLISWLFILNFMPIYIGSNASRGRQRMRHMPNVQSCGGCATCSDYNGCLSKPR 60  
Qy 61 LPPALERIGMKQIGVCLSSCPGTYGTRYPDINKCTKCKADCDTCFKNKFCCKSGPYL 120  
Db 61 LPPALERIGMKQIGVCLSSCPGTYGTRYPDINKCTKCKADCDTCFKNKFCCKSGPYL 120  
Qy 121 HLGKCLDNCPEGLNANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180  
Db 121 HLGKCLDNCPEGLNANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180  
Qy 181 QHPKAGNLCPTNTRKCTVQRKCKGGRGKGRKRRKPKNKGSKEAIPDSKSL 240  
Db 181 QHPKAGNLCPTNTRKCTVQRKCKGGRGKGRKRRKPKNKGSKEAIPDSKSL 240  
Qy 241 SKEIPEQRENKQKKRKKVQDKOKS 265  
Db 241 SKEIPEQRENKQKKRKKVQDKOKS 265

## RESULT 9

US-09-894-912A-16  
; Sequence 16, Application US/09894912A  
; Publication No. US20030044792A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
; FILE REFERENCE: 28110/37260A  
; CURRENT FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: To be assigned  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 16  
; LENGTH: 251  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-894-912A-16

Query Match 92.7%; Score 1400; DB 10; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.3e-106; Indels 0; Gaps 0;  
Matches 251; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0;  
Qy 22 QNASRGRQRMRHMPNVQSCGGCATCSDYNGCLSKPRLPALERIGMKQIGVCLSSCP 81  
Db 1 QNASRGRQRMRHMPNVQSCGGCATCSDYNGCLSKPRLPALERIGMKQIGVCLSSCP 60

QY 82 SGYYGTRYPDINKCTCKKADCDTCFKNKPFCTKCKSGFYHLGKCLDNCPEGLANNHTWE 141  
 DB 61 SGYYGTRYPDINKCTCKKADCDTCFKNKPFCTKCKSGFYHLGKCLDNCPEGLANNHTWE 120  
 QY 142 CVSIVHCVSEVSNWSPCTCKGKTCGPKRGTRVREIIQHPSAKGNLCPPPTNETRKCTV 201  
 DB 121 CVSIVHCVSEVSNWSPCTCKGKTCGPKRGTRVREIIQHPSAKGNLCPPPTNETRKCTV 180  
 QY 202 ORKKCKGKRGKGRKRRKKNKESKEAIPDSKLSLESSKEIPBRENKQKKKKVQD 261  
 DB 181 ORKKCKGKRGKGRKRRKKNKESKEAIPDSKLSLESSKEIPBRENKQKKKKVQD 240  
 QY 262 KOKSVSVSTVH 272  
 DB 241 KOKSVSVSTVH 251

## RESULT 10

US-09-894-912A-32  
 ; Sequence 32, Application US/09894912A  
 ; Publication No. US20030044792A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang et al.  
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
 ; FILE REFERENCE: 28110/37260A  
 ; CURRENT APPLICATION NUMBER: US/09/894,912A  
 ; CURRENT FILING DATE: 2002-05-10  
 ; PRIOR APPLICATION NUMBER: To be assigned  
 ; PRIOR FILING DATE: 2001-04-05  
 ; PRIOR APPLICATION NUMBER: 60/266,614  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: 60/215,733  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: 09/757,562  
 ; PRIOR FILING DATE: 2001-01-09  
 ; PRIOR APPLICATION NUMBER: 09/543,774  
 ; PRIOR FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 32  
 ; LENGTH: 279  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-894-912A-32

Query Match 87.1%; Score 1315.5; DB 10; Length 279;  
 Best Local Similarity 87.1%; Pred. No. 3.7e-99;  
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;  
 QY 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
 DB 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
 QY 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120  
 DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120  
 QY 121 HLKCLDNCPEGLANNHTMECVSIHCVSEVSNWSPCTCKGKTCGPKRGTRVREII 180  
 DB 121 HLKCLDNCPEGLANNHTMECVSIHCVSEVSNWSPCTCKGKTCGPKRGTRVREII 180  
 QY 181 QHPSA--KGNLCPPPTNETRKCTVORKKCKOKGERKKGRKKRKKKPKNGKESKE--AIPDSK 236  
 DB 181 QHPSAKGRGNLCPPPTNETRKCTVORKKCKOKGERKKGRKKRKKKPKNGKESKE--AIPDSK 240  
 QY 237 SLESSKEIPERENK--QKKKKKKVQD--QKSVSVSTVH 272  
 DB 241 GLESSIETPDQENKRRQKKRRARDQKSVSVSTVH 279

## RESULT 11

US-10-185-770-4

; Sequence 4, Application US/10185770  
 ; Publication No. US20030022217A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CECCARDI, Toni et al.  
 ; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; FILE REFERENCE: CL0001247  
 ; CURRENT APPLICATION NUMBER: US/10/185,770  
 ; CURRENT FILING DATE: 2002-07-01  
 ; PRIOR APPLICATION NUMBER: 60/301,852  
 ; PRIOR FILING DATE: 2001-07-02  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 225  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-185-770-4

Query Match 84.8%; Score 1280; DB 14; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-96;  
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
 DB 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
 QY 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120  
 DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120  
 QY 121 HLKCLDNCPEGLANNHTMECVSIHCVSEVSNWSPCTCKGKTCGPKRGTRVREII 180  
 DB 121 HLKCLDNCPEGLANNHTMECVSIHCVSEVSNWSPCTCKGKTCGPKRGTRVREII 180  
 QY 181 QHPSAKGNLCPPPTNETRKCTVORKKCKOKGERKKGRKKRKKPKNK 225  
 DB 181 QHPSAKGNLCPPPTNETRKCTVORKKCKOKGERKKGRKKRKKPKNK 225

## RESULT 12

US-10-087-192-1782  
 ; Sequence 1782, Application US/10087192  
 ; Publication No. US20020182586A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morris, David W.  
 ; APPLICANT: Engelhard, Eric K.  
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
 ; TITLE OF INVENTION: CANCER  
 ; FILE REFERENCE: 529452000122  
 ; CURRENT APPLICATION NUMBER: US/10/087,192  
 ; CURRENT FILING DATE: 2002-03-01  
 ; PRIOR APPLICATION NUMBER: US 09/747,377  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: US 09/798,586  
 ; PRIOR FILING DATE: 2001-03-02  
 ; NUMBER OF SEQ ID NOS: 2059  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1782  
 ; LENGTH: 239  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-087-192-1782

Query Match 80.7%; Score 1218; DB 12; Length 239;  
 Best Local Similarity 97.3%; Pred. No. 2.7e-91;  
 Matches 215; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 60  
 DB 18 MHLRLISWLFILNFMFYIGSNASRGRQRMRHNPVNSQCGGCATCSDYNGCLSKPR 77



```
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 176
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-176

Query Match      73.6%; Score 1111; DB 15; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.1e-82;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GSNASRGRQRMRHPNVSQCGGCATCSYNGCLSKPRLPFALEIRIGMKQIGVCLSS 79
Db 1 GSNASRGRQRMRHPNVSQCGGCATCSYNGCLSKPRLPFALEIRIGMKQIGVCLSS 60

Qy 80 CPSSGYGTRYPDINKCTCKACDCTCFNKNFCTKSGFYHLGKCLDNCPEGLEANNHT 139
Db 61 CPSSGYGTRYPDINKCTCKACDCTCFNKNFCTKSGFYHLGKCLDNCPEGLEANNHT 120

Qy 140 MECVSIHVCEVSENNPWSPTCKGKTGCKRGGTETRVREIIQHPSAKGNLCPTNETRKC 199
Db 121 MECVSIHVCEVSENNPWSPTCKGKTGCKRGGTETRVREIIQHPSAKGNLCPTNETRKC 180

Qy 200 TVQRKKCKQKGERG 212
Db 181 TVQRKKCKQKGERG 193

RESULT 15
US-10-094-886-174
; Sequence 174, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Baha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
```

```
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderna, Steven
; APPLICANT: LaRoche, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 174
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-174
```

```
Query Match      73.3%; Score 1107; DB 15; Length 195;
Best Local Similarity 99.5%; Pred. No. 2.3e-82;
Matches 192; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 20 GSNASRGRQRMRHPNVSQCGGCATCSYNGCLSKPRLPFALEIRIGMKQIGVCLSS 79
Db 1 GSNASRGRQRMRHPNVSQCGGCATCSYNGCLSKPRLPFALEIRIGMKQIGVCLSS 60

Qy 80 CPSSGYGTRYPDINKCTCKACDCTCFNKNFCTKSGFYHLGKCLDNCPEGLEANNHT 139
Db 61 CPSSGYGTRYPDINKCTCKACDCTCFNKNFCTKSGFYHLGKCLDNCPEGLEANNHT 120

Qy 140 MECVSIHVCEVSENNPWSPTCKGKTGCKRGGTETRVREIIQHPSAKGNLCPTNETRKC 199
Db 121 MECVSIHVCEVSENNPWSPTCKGKTGCKRGGTETRVREIIQHPSAKGNLCPTNETRKC 180

Qy 200 TVQRKKCKQKGERG 212
Db 181 TVQRKKCKQKGERG 193
```

Search completed: June 29, 2004, 17:15:58  
Job time : 37.6408 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:56:05 ; Search time 13.8641 Seconds  
(without alignment)  
1887.186 Million cell updates/sec

Title: US-09-894-912A-34  
Perfect score: 1510  
Sequence: 1 MLRLISWLPILNFMRYIG.....QQKKRQVDKQSVSVTVH 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: Pir1.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	12.7	1299	2 T43251	furin (EC 3.4.21.7)
2	186.5	12.4	1548	2 S34583	serine proteinase
3	184	12.2	962	2 JC5571	subtilisin-like pr
4	184	12.2	975	2 JC5570	subtilisin-like pr
5	176.5	11.7	1680	2 A43434	furin (EC 3.4.21.7)
6	173.5	11.5	969	1 A39490	subtilisin-like pr
7	168	11.1	915	2 B48225	probable proprotel
8	167.5	11.1	932	2 F52527	PACE4A - mouse (fr
9	167	11.1	915	1 A48225	subtilisin-like pr
10	162	10.7	899	2 G02428	subtilisin-like pr
11	162	10.7	915	2 G02428	subtilisin-like pr
12	161	10.7	440	2 T24232	hypothetical prote
13	158.5	10.5	379	2 A59180	Wnt inhibitory fac
14	156.5	10.4	937	2 F53282	gene PACE4 protein
15	156	10.3	803	2 A47723	F-spondin precuro
16	153.5	10.2	942	2 D87803	protein bli-4D [im
17	150.5	10.0	1203	2 A49175	Notch B protein -
18	149	9.9	1371	2 A33837	insulin-like growt
19	148.5	9.8	2524	2 A35844	Xotch protein - Af
20	147.5	9.8	2471	2 A49128	cell-rate determin
21	145	9.6	570	2 T37314	probable kexin (EC
22	145	9.6	1620	2 T27283	hypothetical prote
23	143.5	9.5	807	2 A38152	F-spondin - rat
24	142	9.4	378	2 B59180	Wnt inhibitory fac
25	142	9.4	677	2 C42125	trophozoite cystei
26	139	9.2	1367	1 IGHUR1	insulin-like growt
27	139	9.2	2555	2 A40043	notch protein homo
28	138.5	9.2	1369	2 S70713	protein-tyrosine k
29	138	9.1	2531	2 S18108	notch protein homo

30	138	9.1	2531	2 A46019	notch-1 protein -
31	137	9.1	2437	2 S42612	transmembrane prot
32	136	9.0	540	2 B47417	insulin receptor-r
33	135.5	9.0	1382	1 INHUR	insulin receptor p
34	134.5	8.9	1111	2 T26972	hypothetical prote
35	134.5	8.9	1372	2 A34157	insulin receptor p
36	134.5	8.9	1383	2 A36080	insulin receptor p
37	129.5	8.6	2101	2 S57245	insulin receptor (
38	129.5	8.6	2148	1 A56081	insulin receptor -
39	128	8.5	327	2 A46484	apoptosis-mediati
40	128	8.5	1268	2 B36502	insulin receptor-r
41	127	8.4	861	2 A48825	Notch homolog Motc
42	126.5	8.4	3871	2 T22812	hypothetical prote
43	125.5	8.3	667	2 A48579	trophozoite surfac
44	125	8.3	837	2 S43656	furin (EC 3.4.21.7
45	124	8.2	1274	2 T42017	cysteine rich prot

ALIGNMENTS

RESULT 1

T43251

furin (EC 3.4.21.75) - fall armyworm

N:Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; serin

C:Species: Spodoptera frugiperda (fall armyworm)

C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000

C:Accession: T43251

R:Cleplik, M.; Klenk, H.

A:Description: Cloning and functional characterization of FURIN from Spodoptera frugiper

A:Reference number: 222368

A:Accession: T43251

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1299 <CIE>

A:Cross-references: EMBL:Z68888; NID:g1167859; PID:e219690; PIDN:CAA93116.1

A:Experimental source: clone Sfurin 6; ovary

C:Function:

C:Description: responsible for the endoproteolytic processing of proproteins with specif.

C:Keywords: hydrolase; serine proteinase

Query Match	12.7%	Score 192;	DB 2;	Length 1299;
Best Local Similarity	27.8%	Pred. No. 2.7e-05;		
Matches	63;	Conservative	26;	Mismatches 78; Indels 60; Gaps 12;
Qy	36	NVSGCGCGCATCSD-YNGCLSCKPRUFFALERIGMKQIGVCLSCSPSGYGYTRYDPDINK 94		
Db	795	SVCRPCAARCATCERADGCTSCRHLL-----VLHDGTCMASCPSPSHYET---EDDM 843		
Qy	95	CTKCKADCTCF--NKNFCTKCKSGFYHLGLKCLDNCPEGLEANNHTMECVSIHVCEVSE 152		
Db	844	CAKCHSCDTCQGGPBTQCVTCHPSTYALDGRCVTSCPPAYYADKKRKEC---MRCPVG- 899		
Qy	153	WNWPSCTKKGKTCGFKRGTRVRETIQHPSAKGNLCPP---TNETRKC-TVORRKKCK 208		
Db	900	---CSTYT-----SAPCLSCPEKWEINKKCKMPVSGDKCSA 933		
Qy	209	GERKKGRERKRPKNK-----CESKE---AIPDSKSLSSKEIPE 246		
Db	934	GEPAV---DQKCKRNPACSDCYGENEGHCLTCFNPILLQDYKCVPE 977		

RESULT 2

S34583

serine proteinase (EC 3.4.21.-) PC6B - mouse

C:Species: Mus musculus (house mouse)

C>Date: 02-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999

C:Accession: S34583

R:Nakagawa, T.; Murakami, K.; Nakayama, K.

FEBS Lett. 327, 165-171, 1993

A>Title: Identification of an isoform with an extremely large Cys-rich region of PC6, a

A:Reference number: S34583; MUID:93327934; PMID:8335106

A/Accession: S34583  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-1548 <NAK>  
A/Cross-references: GB: D17583; NID: G407344; PIDN: BAA04507.1; PID: G440374  
C/Keywords: hydrolase; serine proteinase

Query Match 12.4%; Score 186.5; DB 2; Length 1548;  
Best Local Similarity 25.0%; Pred. No. 6.9e-05;  
Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;

QY 38 SQCGCGCATCSYNG--CLSKRPLFPALRIGMKQIGVCLSSCPSSGGYGYTRYPD1--N 93  
DB 880 NQCHSSCKTC--NGSLCASCPTGMVLMQ-----ACVPSQPG-----TWPSVTSG 924  
QY 94 KCTKACADCTCNKPNFTCKS-----GFTLHGKCLDNCPEGLNHNHTMB-CVSIIVHC 148  
DB 925 SCEKSEDCVSCGADLCQCLSQPNTLLHREGCVHSCPEGYAKDGVCEHCSS--PC 982  
QY 149 EVSEWNPWSPCTKKG-----KTC-----GFKRGTRVREIIQHPSAKGNL- 189  
DB 983 KTCGNATSCNCEGDFVLHGVCWCTCPKHAVGVGCKHCPERCODCIHEKTKCEMP 1042  
QY 190 -----CPPT--NETRKTCTVQRKCKQKGERGKGRKRKRKPKNKGSKAIPDS 235  
DB 1043 DFLYNDMCHRSCKSPYDMRQCVPHKNC-----LECNPKEDDCKVCADTS 1091  
QY 236 KSLSS--KEIPE-----QRENKO 252  
DB 1092 KALHNGCLDCEPGTKREND 1115

RESULT 3  
JC5571  
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-II -  
C/Species: Homo sapiens (man)  
C/Date: 23-Sep-1997 #sequence\_revision 23-Sep-1997 #text\_change 20-Jun-2000  
C/Accession: JC5571  
R/Mori, K.; Kii, S.; Tsuji, A.; Nagahama, M.; Imamaki, A.; Hayashi, K.; Akamatsu, T.; Na  
J. Biochem. 121, 941-948, 1997  
A/Title: A novel human PACE4 isoform, PACE4E is an active processing protease containing  
A/Reference number: JC5570; MUID: 97335942; PMID: 9192737  
A/Accession: JC5571  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-962 <MOR>  
A/Cross-references: DDBJ: D87994; NID: G2330550; PIDN: BAA21792.1; PID: G2330551  
A/Experimental source: brain cerebellum  
C/Comment: This enzyme is a processing protease and responsible for processing of various  
ch it is retained intracellularly.

Query Match 12.2%; Score 184; DB 2; Length 962;  
Best Local Similarity 26.6%; Pred. No. 7.1e-05;  
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;

QY 18 YIGSNASRRRQRRMHPNVSCQCGCATCS--YNGCLSKRPLFPALRIGMKQIGV 75  
DB 724 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 766  
QY 76 CLSSCPSSGYGYTRYPDINKCKACDCTCNK-NFTCKCKSGFYHLGKCLDNCPEGL 134  
DB 76 CLSSCPSSGYGYTRYPDINKCKACDCTCNK-NFTCKCKSGFYHLGKCLDNCPEGL 134

RESULT 4  
JC5570  
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-I -  
C/Species: Homo sapiens (man)  
C/Date: 23-Sep-1997 #sequence\_revision 23-Sep-1997 #text\_change 20-Jun-2000  
C/Accession: JC5570  
R/Mori, K.; Kii, S.; Tsuji, A.; Nagahama, M.; Imamaki, A.; Hayashi, K.; Akamatsu, T.; Na  
J. Biochem. 121, 941-948, 1997  
A/Title: A novel human PACE4 isoform, PACE4E is an active processing protease containing  
A/Reference number: JC5570; MUID: 97335942; PMID: 9192737  
A/Accession: JC5570  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-975 <MOR>  
A/Cross-references: DDBJ: D87993; NID: G2330548; PIDN: BAA21791.1; PID: G2330549  
A/Experimental source: brain cerebellum  
C/Comment: This enzyme is a processing protease and responsible for processing of various  
ch it is retained intracellularly.

Query Match 12.2%; Score 184; DB 2; Length 975;  
Best Local Similarity 26.6%; Pred. No. 7.2e-05;  
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;

QY 18 YIGSNASRRRQRRMHPNVSCQCGCATCS--YNGCLSKRPLFPALRIGMKQIGV 75  
DB 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779  
QY 76 CLSSCPSSGYGYTRYPDINKCKACDCTCNK-NFTCKCKSGFYHLGKCLDNCPEGL 134  
DB 780 CVTLCPAGFYADE--SQKNCLKCHPSCKKCVDEPKCTCKGFSLARGSCIPCEPGTY 837  
QY 135 ANNHTMEC-----VSIVHCVS--EWNPSPTCKGKTRGRTETRVRE 178  
DB 838 FDSLRRCGECHTCTGTCVGPGRBECIHCAKNFPHDM-----KCVPCAGSGFYPE----- 888  
QY 179 IIGHPSAKGNLC-----PPTNETRKTCTVQRKCKQKGERGKGRKRKRKPKNKG 227  
DB 889 --EMPGLPHKVCRRYGPAGE-RQATVS-----SKGVPG--GQSLASASSPAGE 932

RESULT 5  
A43434  
furin (EC 3.4.21.75) 2 - fruit fly (Drosophila melanogaster)  
C/Species: Drosophila melanogaster  
C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Mar-2000  
C/Accession: A43434  
R/Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dumke, U.; Rentrop, M.; Gateff, E.  
J. Biol. Chem. 267, 17208-17215, 1992  
A/Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc  
A/Reference number: A43434; MUID: 92381036; PMID: 1512259

A:Accession: A43434  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1680 <ROB>  
A:Cross-references: GB:M94375; NID:g157461; PID:g157462  
A:Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBI:P:111934)  
C:Genetics:  
A:Gene: FlyBase:Fur2  
A:Cross-references: FlyBase:FBgn0004598  
C:Superfamily: subtilisin homology  
C:Keywords: hydrolase; serine proteinase; transmembrane protein  
P:409-652/Domain: subtilisin homology <SBT>  
P:418,457,638/Active site: Asp, His, Ser #status predicted

Query Match 11.7%; Score 176.5; DB 2; Length 1680;  
Best Local Similarity 28.0%; Pred. No. 0.00033;  
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

QY 2 HRLISWLFILNFMEXYIGSONASRRQRMRHNPVSGCGGCATCSDY-NGCLSKKPR 60  
DB 1038 HLRVID-LAVCLQPCPDGYFENS-----NRTYCV-----CBPNCASQDHPYCTSCDHH 1087  
QY 61 LPPALERIGMKQIGVGLSSCPSPGYGTRYPDINKTKCKADCDTCF--NKNFCTKCKSGF 118  
DB 1088 LVMEHK-----CYSACPLDTYET---EDNKCAPCHSTCATCGPTDQDCITCRSSR 1136  
QY 119 YLHLGKCLDNCPEGLAANNHTMBCVSIHVCEVSEWNPSPCTKGTGCFRGTETRV-R 177  
DB 1137 YAWQNKCLISCPDGFYADKKRLQCM-----PQEGCKTC-----TSNGVCS 1177  
QY 178 BIIQHPGAKGNLCPPTNETRKTQVRKK-CQKGE 210  
DB 1178 ECLQWT-----LNKRDKIVSGSGCSESE 1203

RESULT 6  
A39490  
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form A - human  
A:Alternate names: kexin homology  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Mar-2000  
C:Accession: A39490  
R:Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landsberg, K.B.; Saltman, D.; Barr, P.J.  
DNA Cell Biol. 10, 757-769, 1991  
A:Title: Identification of a second human subtilisin-like protease gene in the fes/fps  
A:Reference number: A39490; MUID:92075167; PMID:1741956  
A:Accession: A39490  
A:Molecule type: mRNA  
A:Residues: 1-969 <KIE>  
A:Cross-references: GB:M80482; NID:g189531; PIDN:AAA59998.1; PID:g189532  
C:Genetics:  
A:Gene: GDB:PACE4  
A:Cross-references: GDB:131390; OMIM:167405  
A:Map position: 15q26-15q26  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: alternative splicing; hydrolase; serine proteinase  
P:150-969/Product: serine proteinase PACE4 #status predicted <SIG>  
P:196-434/Domain: subtilisin homology <SBT>  
P:205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 11.5%; Score 173.5; DB 1; Length 969;  
Best Local Similarity 28.8%; Pred. No. 0.00035;  
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 18 YIGSONASRRQRMRHNPVSGCGGCATCSDY-NGCLSKKPRLPFALERIGMKQIGV 75  
DB 737 YFGDTAARRCR-----CHKGCTCSSRAATCLSCR-RGFY-----HHQENWT 779  
QY 76 CLSSCPSPGYGTRYPDINKTKCKADCDTCFNK-NFTYKCKSGFYLHLGKCLDNCPEGL 134  
DB 780 CVTLCFAGFYADE--SQKNCCLKPCKCKKCVDEPEKCTVCKEFGSLARGSCIPDCEPGY 837  
QY 135 ANNHTMBCVSIHVCEVSEWNPSPCTKGTGCTG 167

DB 838 FDSELIIRGRCCH-----TCG 853  
RESULT 7  
B48225  
probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat  
A:Alternate names: PCS precursor  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 23-Feb-1997  
C:Accession: B48225  
R:Lusson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993  
A:Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate pro-  
A:Reference number: A48225; MUID:93342056; PMID:8341687  
A:Accession: B48225  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-915 <LUS>  
A:Cross-references: GB:L14933  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
C:Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase  
P:1-34/Domain: signal sequence #status predicted <SIG>  
P:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>  
P:117-915/Product: probable proprotein convertase 5 #status experimental <MAT>  
P:164-402/Domain: subtilisin homology <SBT>  
P:173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.1%; Score 168; DB 2; Length 915;  
Best Local Similarity 26.3%; Pred. No. 0.00076;  
Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;

QY 35 PNVSQ-GCQG-----GCATCSYNGCLSKKPRLPFALERIGMKQIGVGLSSCPSPGYGTRY 89  
DB 640 PECSEVCGDGPDPDHTCDLHYHYLKNTR-----ICVSSCPGHP---H 682  
QY 90 PDINKTKCKADCDTCFNKPN--CTKCKSGFYLH--LGKCLDNCPEGLAANNHTMBCVSI 145  
DB 683 ADKKCRKCAVNCSSCFSGSHADQCLSKYGYFLNBTSSVACQPEGSVDIKGNIC--- 739  
QY 146 VHCSEWNPSPCTKGTGCTGCFRGTETRVREIIQHPHSAKGNLCPPTNETRKTQVRK 204  
DB 740 -----GKSENCKTCTGPHNCTE-----CKGGL---SLQGRCSV--- 771  
QY 205 KQKGE 210  
DB 772 TCEDQG 777

RESULT 8  
152527  
PACE4A - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999  
C:Accession: 152527  
R:Hosaka, M.; Murakami, K.; Nakayama, K.  
Biomed. Res. 15, 383-390, 1994  
A:Title: PACE4A is a ubiquitous endoprotease that has similar but not identical substrat-  
A:Reference number: 152527  
A:Accession: 152527  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-932 <RES>  
A:Cross-references: GB:D50060; NID:g769700; PIDN:BAJ08777.1; PID:g769701  
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
P:172-410/Domain: subtilisin homology <SBT>

Query Match 11.1%; Score 167.5; DB 2; Length 932;  
Best Local Similarity 24.1%; Pred. No. 0.00083;  
Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

QY 18 YIGSONASRRQRMRHNPVSGCGGCATCSDY-NGCLSKKPRLPFALERIGMKQIGV 75







F-spondin precursor - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 27-Jun-1994 #sequence-revision 27-Jun-1994 #text\_change 17-Nov-2000  
C:Accession: A47723  
R:Ruiz i Altaba, A.; Cox, C.; Jessell, T.M.; Klar, A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 8268-8272, 1993  
A:Title: Ectopic neural expression of a floor plate marker in frog embryos injected with  
A:Reference number: A47723; MUID:93376785; PMID:8367492  
A:Accession: A47723  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-803 <RUI>  
A:Cross-references: GB:I09123; NID:9409244; PIDN:AAA19105.1; PID:9409245  
C:Superfamily: F-spondin; thrombospondin type 1 repeat homology  
F:435-489/Domain: thrombospondin type 1 repeat homology <THR2>  
F:607-662/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 10.3%; Score 156; DB 2; Length 803;  
Best Local Similarity 24.6%; Pred. No. 0.0042;  
Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;

Qy 37 VSQGGGGGATCSDYNGCLSCCKPRLFPALBRIGMKQIGVCLSSCPSS--GYGTTRYPDINK 94  
Db 544 VNEECSPSSCIVTEWAEWECSS-----ATCRMGKKRHRMIKMTPADGSMCKADYITEVEK 598  
Qy 95 C--TKCK-----ADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLEANN--- 137  
Db 599 CMPECHTTPCVLSPWSENDCSVTCGKGTTRTQR---MLKSPSELGDCNEBELKQVEK 655  
Qy 138 -HTMECVSIVHCEVSEWNPSPCTKKGKTCGFKRGTTETVRBIIQHPSAKGNLCPTNET 196  
Db 656 CMLPECP--ISCELTWYSWSEC--NKSCG--KGMIRTRMITMBEPQFGAVCPETVQR 708  
Qy 197 RKCTVQRKCKQGERKKGRKKRKPKNKGSKEAIPDSKSLSSKEIPEQRENKQ 252  
Db 709 KKCRLL--RKQCK-----SSGNERRHLK-----DAREKRSEKIKEDSDGBQ 747

Search completed: June 29, 2004, 17:03:06  
Job time : 13.8641 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:24 ; Search time 8.25243 Seconds  
(without alignments)  
1716.232 Million cell updates/sec

Title: US-09-894-912A-34

Perfect score: 1510

Sequence: 1 MHLRLISLWFLIILNFMVEYG.....QQKKRKVDKQKSVSVSTVH 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	12.4	1877	1 PCK5 MOUSE	Q04592 mus musculus
2	184.5	12.2	1696	1 PCK5 BRACL	Q9nj15 branchiostoma
3	176.5	11.7	1679	1 PUR2 DROME	P30432 drosophila
4	173.5	11.5	969	1 PAC4 HUMAN	P29122 homo sapien
5	168	11.1	1877	1 PCK5 RAT	P41413 rattus norv
6	162	10.7	913	1 PCK5 HUMAN	Q92824 homo sapien
7	158.5	10.5	379	1 WIF1 HUMAN	Q9y5w5 homo sapien
8	157.5	10.4	374	1 WIF1 XENLA	Q9w6f8 xenopus lae
9	156.5	10.4	937	1 PAC4 RAT	Q63415 rattus norv
10	156	10.3	803	1 PSPO XENLA	P35447 xenopus lae
11	150.5	10.0	2470	1 NTC2 MOUSE	Q35516 mus musculus
12	149	9.9	1370	1 IGIR RAT	P24062 rattus norv
13	148.5	9.8	2524	1 NOTC XENLA	P21783 xenopus lae
14	147.5	9.8	2471	1 NTC2 RAT	Q9qw30 rattus norv
15	145	9.6	943	1 BLI4 CARBL	P51559 caenorhabdi
16	143.5	9.5	379	1 WIF1 MOUSE	Q9wual mus musculus
17	143.5	9.5	807	1 PSPO RAT	P35446 rattus norv
18	142	9.4	378	1 WIF1 BRARE	Q9w6f9 brachydanio
19	142	9.4	1373	1 IGIR MOUSE	Q60751 mus musculus
20	139	9.2	1367	1 IGIR HUMAN	P08069 homo sapien
21	139	9.2	2556	1 NTC1 HUMAN	P46531 homo sapien
22	138	9.1	2531	1 NTC1 MOUSE	Q01705 mus musculus
23	138	9.1	2531	1 NTC1 RAT	Q07008 rattus norv
24	137	9.1	2437	1 NTC1 BRARE	P46530 brachydanio
25	135.5	9.0	1382	1 INSR HUMAN	P06213 homo sapien
26	134.5	8.9	1372	1 INSR MOUSE	P15208 mus musculus
27	134.5	8.9	1383	1 INSR RAT	P15127 rattus norv
28	133.5	8.8	2482	1 VWF FIG	Q28833 sus scrofa
29	130.5	8.6	1300	1 IRR MOUSE	Q9wcl4 mus musculus
30	130	8.6	2471	1 NTC2 HUMAN	Q04721 homo sapien
31	129.5	8.6	2146	1 INSR DROME	P09208 drosophila
32	128	8.5	327	1 TNR6 MOUSE	P25446 mus musculus
33	128	8.5	1297	1 IRR HUMAN	P14616 homo sapien

34 126 8.3 687 1 VS41 GIALA  
35 125.5 8.3 667 1 TS11 GIALA  
36 123 8.1 461 1 TRIA\_PIG  
37 123 8.1 559 1 TRAP\_PLAFA  
38 123 8.1 581 1 IRR\_RAT  
39 119.5 7.9 1429 1 LI12 CARBL  
40 119 7.9 1300 1 IRR\_CAVPO  
41 119 7.9 2703 1 NOTC DROME  
42 118.5 7.8 1477 1 HTK7 HYDAT  
43 118 7.8 867 1 SSPO\_BOVIN  
44 117.5 7.8 833 1 DL\_DROME  
45 117.5 7.8 1367 1 LT23\_CARBL

#### ALIGNMENTS

RESULT 1  
PCK5 MOUSE  
ID PCK5 MOUSE STANDARD; PRT; 1877 AA.  
AC Q04592; Q62040;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Proprotein convertase subtilisin/kexin type 5 precursor (BC 3.4.21.-)  
DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PCS) (PC6)  
DE (Subtilisin-like proprotein convertase 6) (SPC6).  
GN PCK5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PCSB).  
RC STRAIN=ICR; TISSUE=Intestine;  
RX MEDLINE=93327934; PubMed=8335106;  
RA Nakagawa T., Murakami K., Nakayama K.;  
RT "Identification of an isoform with an extremely large Cys-rich region of PCS, a Kex2-like processing endoprotease.";  
RT FEBS Lett. 327:165-171(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM PCSA).  
RC TISSUE=Brain, and Intestine;  
RX MEDLINE=93324489; PubMed=8468318;  
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;  
RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACR4.";  
RT J. Biochem. 113:132-135(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM PCSA).  
RC TISSUE=Adrenal cortex;  
RX MEDLINE=93342056; PubMed=8341687;  
RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";  
RT Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
RN [4]  
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
RC MEDLINE=97103178; PubMed=8947550;  
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;  
RT "The isoforms of proprotein convertase PCS are sorted to different subcellular compartments.";  
RT J. Cell Biol. 135:1261-1275(1996).  
RN [5]  
RP DEVELOPMENTAL EXPRESSION.  
RC MEDLINE=96293359; PubMed=8698813;  
RA Constam D.B., Calton M., Robertson B.J.;  
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

P92127 giardia lam  
Q03185 giardia lam  
P50555 sus scrofa  
P16893 plasmodium  
Q64716 rattus norv  
P14585 caenorhabdi  
P14617 cavia porce  
P07207 drosophila  
Q25197 hydra atten  
P98167 bos taurus  
P10041 drosophila  
P24348 caenorhabdi





OC Ephydroidea: Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R, Tuebingen, and Iso-1;  
 RX MEDLINE=92381036; PubMed=1512259;  
 RA Roebroeck A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,  
 RA Rentrop M., Gateff E.A.P., Leunissen J.A.M., van de Ven W.J.M.;  
 RT "Cloning and functional expression of Dfurin2, a subtilisin-like  
 RT protein processing enzyme of Drosophila melanogaster with multiple  
 RT repeats of a cysteine motif.";  
 RL J. Biol. Chem. 267:17208-17215(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Iso-1;  
 RX MEDLINE=95186060; PubMed=7880443;  
 RA Roebroeck A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,  
 RA van de Ven W.J.M.;  
 RT "The Dfur2 gene of Drosophila melanogaster: genetic organization,  
 RT expression during embryogenesis, and pro-protein processing activity  
 RT of its translational product Dfurin2.";  
 RL DNA Cell Biol. 14:223-234 (1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baillew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies C.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Finkler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -!- FUNCTION: Furin is likely to represent the ubiquitous endoprotease  
 CC activity within constitutive secretory pathways and capable of  
 CC cleavage at the RX(K/R)R consensus motif (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC proproteins by cleavage of Arg-Xaa-Yaa-Arg|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,  
 CC complement component C3 and von Willebrand factor from their  
 CC respective precursors.  
 CC -!- TISSUE SPECIFICITY: Transient expression in a subset of central

nervous system neurons during embryonic stages 12-13. Expression  
 in developing tracheal tree from stage 13 to end of embryonic  
 development.  
 -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.  
 -!- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; M943375; AAA28551.1; --  
 EMBL; L33831; AAA69860.1; --  
 EMBL; AB003502; AAF48598.1; --  
 FIR; A43434; A43434.  
 HSP; Q99405; IMPT.  
 MEROPS; S08.049; --  
 FlyBase; FBgn0004598; Fur2.  
 GO; GO:0004276; P: furin activity; IDA.  
 InterPro; IPR006212; Furin repeat.  
 InterPro; IPR009030; Grow\_fac\_recep.  
 InterPro; IPR000209; Peptidase\_S8.  
 InterPro; IPR002884; Peptidase\_S8B.  
 InterPro; IPR009020; Protease\_inhib.  
 Pfam; PF01493; P\_proteasein; 2.  
 Pfam; PF00082; Peptidase\_S8; 1.  
 PRINTS; PR00723; SUBTILISIN.  
 ProDom; PD000717; P\_domain; 1.  
 SMART; SM00261; FU; 10.  
 PROSITE; PS00136; SUBTILASE ASP; 1.  
 PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KX Hydroxylase; Serine protease; Glycoprotein; Signal; Transmembrane;  
 KW Multigene family; Zymogen; Repeat.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT PROPEP 2 318 POTENTIAL.  
 FT CHAIN 319 1679 FURIN-LIKE PROTEASE 2.  
 FT ACT\_SITE 417 417 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 456 456 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DOMAIN 961 1443 10 X TANDEM REPEATS, CYS-RICH.  
 FT REPEAT 961 1006 1.  
 FT REPEAT 1007 1056 2.  
 FT REPEAT 1057 1103 3.  
 FT REPEAT 1104 1152 4.  
 FT REPEAT 1153 1204 5.  
 FT REPEAT 1205 1253 6.  
 FT REPEAT 1254 1298 7.  
 FT REPEAT 1299 1345 8.  
 FT REPEAT 1346 1392 9.  
 FT REPEAT 1393 1443 10.  
 FT TRANSMEM 1512 1532 POTENTIAL.  
 FT DOMAIN 1533 1679 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 3 3 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 205 205 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 442 442 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 480 480 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 927 927 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 1060 1060 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 1181 1181 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 1439 1439 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CONFLICT 152 153 MISSING (IN REF. 1).  
 FT CONFLICT 177 177 V -> F (IN REF. 1).  
 FT CONFLICT 213 213 V -> VDQL (IN REF. 1).  
 SQ SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CF6 CRC64;



Query Match 11.74; Score 176.5; DB 1; Length 1679;  
 Best Local Similarity 28.04; Pred. No. 2.3e-05;  
 Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

QY 2 HLRISLWFLIINPMYBIGSONASRGRQRMRHPNVSGCGGCATCSDY-NGCLSKCKPR 60  
 DB 1037 HLAVID-LAVCLQCPDGPENP-----RNRTCVF-----CEPNCASQDHPYCTSCDHH 1086

QY 61 LPFALEIRIGHKQIGVCLSSCPGSGYGYTRYDINKTKTKADCTCP--NNFCTKCKSGF 118  
 DB 1087 LVNHEHK-----CYSACPDLDTYET---EDNRCAPCHSTCATCGPTDQDCTICSSR 1135

QY 119 YLHLGKCLDNCPEGLANNHMECVSVHCVSEWNPSPCTKGGKTCGPKRGKTETRV-R 177  
 DB 1136 YANQNKCLISCPDGFYADKKKLECH-----PCQEGCKTC-----TSNGVCS 1176

QY 178 EIIQHPKAGNLCPTTNETRCKTVQRK-CQKGE 210  
 DB 1177 ECLQNTI-----LNKRDKCIVSGEGCSESE 1202

RESULT 4

PAC4 HUMAN STANDARD; PRT; 969 AA.  
 ID PAC4 HUMAN STANDARD; PRT; 969 AA.  
 AC P29122; Q15099; Q15100; Q9UEJ7; Q9UEJ2; Q9UEJ3; Q9UEJ8;  
 AC Q9UEJ9; Q9Y4G9; Q9Y4H0; Q9Y4H1;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DB Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)  
 DB (Subtilisin/kexin-like protease PACB4) (Subtilisin-like proprotein  
 DB convertase 4) (SPC4).  
 GN PACB4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS PACB4A-I AND PACB4B).  
 RC TISSUE=Hepatoma, and Kidney;  
 RX MEDLINE=92075167; PubMed=1741956;  
 RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,  
 RA Barr P.J.;  
 RT Identification of a second human subtilisin-like protease gene in  
 RT the tes/fps region of chromosome 15;  
 RL DNA Cell Biol. 10:757-769(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS PACB4C AND PACB4D).  
 RC TISSUE=Placenta;  
 RX MEDLINE=94235049; PubMed=8179631;  
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,  
 RA Matsuda Y.;  
 RT Identification of novel cDNAs encoding human kexin-like protease,  
 RT PACB4 isoforms.;  
 RL Biochem. Biophys. Res. Commun. 200:943-950(1994).  
 RN [3]  
 RP ERRATUM.  
 RX MEDLINE=95071480; PubMed=7980617;  
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,  
 RA Matsuda Y.;  
 RT Identification of novel cDNAs encoding human kexin-like protease,  
 RT PACB4 isoforms.;  
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM PACB4A-II).  
 RC TISSUE=Placenta;  
 RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,  
 RA Matsuda Y.;  
 RT Identification of a novel PACB4 isoform, PACB4B.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS PACB4E-I AND PACB4E-II).  
 RC TISSUE=Cerebellum;

RA MEDLINE=97335942; PubMed=9192737;  
 RX Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,  
 RA Akamatsu T., Nagamune H., Matsuda Y.;  
 RT "A novel human PACB4 isoform, PACB4B is an active processing protease  
 RT containing a hydrophobic cluster at the carboxy terminus.";  
 RL J. Biochem. 121:941-948(1997).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORMS PACB4A-I; PACB4A-II; PACB4CS; PACB4D;  
 RP PACB4E-I AND PACB4E-II).  
 RX MEDLINE=96021085; PubMed=9378725;  
 RA Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,  
 RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;  
 RT "Genomic organization and alternative splicing of human PACB4 (SPC4),  
 RT kexin-like processing endoprotease.";  
 RL J. Biochem. 122:438-452(1997).  
 RN [7]  
 RP ALTERNATIVE SPLICING (ISOFORM PACB4CS).  
 RX MEDLINE=97064242; PubMed=8906861;  
 RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;  
 RT "Functional analysis of human PACB4-A and PACB4-C isoforms:  
 RT identification of a new PACB4-CS isoform.";  
 RL FEBS Lett. 396:31-36(1996).  
 RN [8]  
 RP CHARACTERIZATION.  
 RX MEDLINE=99233559; PubMed=10215603;  
 RA Sucic J.P., Moehring J.M., Innocencio N.M., Luchini J.W.,  
 RA Moehring T.J.;  
 RT "Endoprotease PACB4 is Ca2+-dependent and temperature-sensitive and  
 RT can partly rescue the phenotype of a furin-deficient cell strain.";  
 RL Biochem. J. 339:639-647(1999).  
 RN [9]  
 RP PROCESSING.  
 RX MEDLINE=96408849; PubMed=9738469;  
 RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,  
 RA Tsuji A., Matsuda Y.;  
 RT "Biosynthetic processing and quaternary interactions of proprotein  
 RT convertase SPC4 (PACB4).";  
 RL FEBS Lett. 434:155-159(1998).  
 CC -1- FUNCTION: Likely to represent an endoprotease activity within the  
 CC constitutive secretory pathway, with unique restricted  
 CC distribution in both neuroendocrine and non-neuroendocrine tissues  
 CC and capable of cleavage at the RK(R/R)R consensus motif.  
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -1- COFACTOR: Calcium (Potential).  
 CC -1- SUBUNIT: The PACB4A-I precursor protein seems to exist in the  
 CC reticulum endoplasmic as both a monomer and a dimer-sized complex  
 CC whereas mature PACB4A-I exists only as a monomer, suggesting that  
 CC propeptide cleavage affects its tertiary or quaternary structure.  
 CC -1- SUBCELLULAR LOCATION: PACB4A-I and PACB4A-II are secreted. PACB4C  
 CC and PACB4CS are not secreted and remain probably in zymogen form  
 CC in endoplasmic reticulum. PACB4E-I and PACB4E-II are retained  
 CC intracellularly probably through a hydrophobic cluster in their C-  
 CC terminus. PACB4B might be secreted.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Rvent=Alternative splicing; Named isoforms=8;  
 CC Name=PACB4A-I; Synonyms=PACB4;  
 CC IsoId=P29122-1; Sequence=Displayed;  
 CC Name=PACB4A-II;  
 CC IsoId=P29122-2; Sequence=VSP\_005436;  
 CC Name=PACB4B; Synonyms=PACB4.1;  
 CC IsoId=P29122-3; Sequence=VSP\_005428, VSP\_005429;  
 CC Note=Probably enzymatically inactive;  
 CC Name=PACB4C;  
 CC IsoId=P29122-4; Sequence=VSP\_005432, VSP\_005433;  
 CC Note=Probably enzymatically inactive;  
 CC Name=PACB4CS;  
 CC IsoId=P29122-5; Sequence=VSP\_005430, VSP\_005431;  
 CC Note=Probably enzymatically inactive;  
 CC Name=PACB4D;  
 CC IsoId=P29122-6; Sequence=VSP\_005427, VSP\_005434, VSP\_005435;  
 CC Note=Probably enzymatically inactive;



proteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys.

-1- SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH EARLY ENDOSOMES (BY SIMILARITY).

-1- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;  
Comment=Additional isoforms seem to exist;

Names=PCSB; Synonyms=Long;  
IsoId=P41413-1; Sequence=displayed;  
Name=PCSA; Synonyms=Short;  
IsoId=P41413-2; Sequence=VSP 005440; VSP 005441;

-1- TISSUE SPECIFICITY: Expressed in the intestine, brain, adrenal gland, anterior pituitary, thyroid, ovaries, testis and lung. Highest levels are found in the gut, duodenum, jejunum and ileum. Expression is higher in female than in male reproductive organs.

-1- DEVELOPMENTAL STAGE: First detected at E9 in highly restricted regions of the neural tube, in caudal myotomes, and at the materno-embryonic junction of the uterus. At E10, restricted expression is detected in the optic and otic vesicles, the roof of midbrain, and trunk myotomes. By midgestation (E13-E16), expression in the developing nervous system has expanded to multiple regions including hippocampus, thalamus, hypothalamus, brain stem, and spinal cord. Expression is also detected in several peripheral organ systems, including gut, lung, adrenal and kidney primordia.

-1- DOMAIN: The propeptide domain acts as an intramolecular chaperone assisting the folding of the zymogen within the endoplasmic reticulum.

-1- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain sorting information. AC 1 directs TGN localization and interacts with the TGN sorting protein PACS-1.

-1- SIMILARITY: Belongs to peptidase family S8.

-1- SIMILARITY: Contains 1 homo B/P domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; L14933; AAA99906.1; --  
EMBL; U47014; AAA87888.1; --  
PIR; B48225; B48225.  
HSP; Q99405; 1MPT.  
MEROPS; S08.076; --  
InterPro; IPR006212; Furin repeat.  
InterPro; IPR000209; Peptidase\_S8.  
InterPro; IPR002884; Peptidase\_S88.  
InterPro; IPR009020; Protease\_inhib.  
Pfam; PF01483; P\_protease; 1.  
PRINTS; PR00723; SUBTILISIN.  
ProDom; PD000717; P\_domain; 1.  
SMART; SM00261; FU; 6.  
PROSITE; PS00136; SUBTILASE\_ASP; 1.  
PROSITE; PS00137; SUBTILASE\_HIS; 1.  
PROSITE; PS00138; SUBTILASE\_SER; 1.  
Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
Cleavage on pair of basic residues; Alternative splicing; Repeat;  
Transmembrane.  
SIGNAL 1 34 BY SIMILARITY.  
PROPEP 35 116 BY SIMILARITY.  
CHAIN 117 1877 PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5.  
DOMAIN 117 1768 EXTRACELLULAR (POTENTIAL).  
TRANSMEM 1769 1789 POTENTIAL.  
DOMAIN 1790 1877 CYTOPLASMIC (POTENTIAL).  
DOMAIN 117 452 CATALYTIC.  
DOMAIN 464 602 HOMO B.  
DOMAIN 638 1753 CYS-RICH MOTIF (CRM) REGION.

DOMAIN 1825 1844 AC 1.  
DOMAIN 1856 1877 AC 2.  
SITE 116 117 CLEAVAGE (AUTO-) (BY SIMILARITY).  
SITE 521 523 CELL ATTACHMENT SITE (POTENTIAL).  
ACT SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).  
ACT SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).  
ACT SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).  
CARBOHYD 227 227 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 383 383 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 667 667 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 754 754 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 804 804 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 854 854 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 1710 1710 N-LINKED (GLCNAC) (POTENTIAL).  
CARBOHYD 1732 1732 N-LINKED (GLCNAC) (POTENTIAL).  
VARSPILIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX -->  
ATEESWAGGFCMLVKKNNLCQKRVLQQLCKKCTCFQG  
(in isoform PCSA).  
/FTid=VSP 005440.  
FT VARSPILIC 916 1877 Missing (in isoform PCSA).  
FT VARSPILIC 916 1877 Missing (in isoform PCSA).  
SQ SEQUENCE 1877 AA; 207888 MW; 890955DC60534444 CRC64;  
Query Match 11.1%; Score 168; DB 1; Length 1877;  
Best Local Similarity 26.3%; Pred. No. 0.0001;  
Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;  
QY 35 PNVSQ-GCOG-----GCATCDYNGCLSCKPLFPALERIGMKQIGVCLSSCPGYYGTRY 89  
DB 640 PECSEVCGDGPDHCTDCLHYHYKLNTR-----ICVSSCPGPHF---H 682  
QY 90 PDINKCTKCADCTCPKNP--CTKCKSGPYLH--LGKCLDNCPEGLAANNHTMBCVSI 145  
DB 683 ADKCRKCAPNCSCPGSHADQLCKYGFLEETSSVCAQPEGSYQDKKNIC--- 739  
QY 146 VHCSESWNPSPCTKKKTC-GFKRGTETRVREIIQHPSAKGNLCPTNETRCKTVORK 204  
DB 740 -----GKSENCKTCTGFHNCTE-----CKGGL---SLQSSRCVS--- 771  
QY 205 KCKQGE 210  
DB 772 TCEDGQ 777  
RESULT 6  
PKCS\_HUMAN STANDARD; PRT; 913 AA.  
ID PKCS\_HUMAN  
AC Q92824; Q13527;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 15-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Proprotein convertase subtilisin/kexin type 5 precursor (BC 3.4.21.--)  
DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PCS) (PC6)  
DE (hPC6).  
GN PKCS OR PCS OR PC6.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=T-cell;  
RX MEDLINE=96353890; PubMed=8755538;  
RA Miranda L., Wolf J., Pichuanes S., Duke R., Franzosooff A.;  
RT "Isolation of the human PC6 gene encoding the putative host protease  
for HIV-1 gp160 processing in CD4+ T lymphocytes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700(1996).  
RN [2]  
RP REVISIONS.  
RA Franzosooff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 15-913 FROM N.A.

RA Reudelhuber T.L.; to the EMBL/GenBank/DBJ databases.  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: Likely to represent a widespread endoprotease activity  
 CC within the constitutive and regulated secretory pathway. Capable  
 CC of cleavage at the RX(K/R)R consensus motif.  
 CC -I- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -I- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -I- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=1;  
 CC Comment=2 isoforms may be produced;  
 CC Name=1;  
 CC IsoId=Q92824-1; Sequence=Displayed;  
 CC TISSUE SPECIFICITY: Expressed in T-lymphocytes.  
 CC -I- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
 CC assisting the folding of the zymogen within the endoplasmic  
 CC reticulum.  
 CC -I- SIMILARITY: Belongs to peptidase family S8.  
 CC -I- SIMILARITY: Contains 1 homo B/P domain.  
 CC -I- SIMILARITY: Contains 1 PLAC domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U56387; AAC50643.2; -;  
 CC EMBL; U49114; AAA91807.1; -;  
 CC PIR; G02428; G02428.  
 CC HSP; Q99405; IMPT.  
 CC MEROPS; S08.076; -;  
 CC Genew; HGNC:8747; PCSK5.  
 CC MIT; 600488; -;  
 CC GO; GO:0005615; C:extracellular space; TAS.  
 CC GO; GO:0004252; F:serine-type endopeptidase activity; TAS.  
 CC GO; GO:0007267; P:cell-cell signaling; TAS.  
 CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
 CC InterPro; IPR006212; Furin repeat.  
 CC InterPro; IPR000209; Peptidase S8.  
 CC InterPro; IPR002884; Peptidase\_S8B.  
 CC InterPro; IPR009020; Protease\_inhib.  
 CC Pfam; PF01483; P:protease; 1.  
 CC Pfam; PF00082; Peptidase S8; 1.  
 CC PRINTS; PR00723; SUBTILISIN.  
 CC ProDom; PD000717; P domain; 1.  
 CC SMART; SM00261; FU\_5.  
 CC PROSITE; PSS0900; PLAC; 1.  
 CC PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 CC PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 CC PROSITE; PS00138; SUBTILASE\_SER; 1.  
 CC Hydrolase; Serine protease; Glycoprotein; Zymogen; signal;  
 CC Cleavage on pair of basic residues; Repeat; Alternative splicing.  
 CC SIGNAL 1 32  
 CC BY SIMILARITY.  
 CC PROPEP 33 114  
 CC CHAIN 115 913  
 CC TYPE 5.  
 CC PROTEIN CONVERTASE SUBTILISIN/KEXIN  
 CC CATALYTIC.  
 CC HOMO B.  
 CC CYS-RICH MOTIF (CRM) REGION.  
 CC PLAC.  
 CC CLEAVAGE (AUTO-) (BY SIMILARITY).  
 CC CELL ATTACHMENT SITE (POTENTIAL).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 115 454  
 CC DOMAIN 462 600  
 CC DOMAIN 636 868  
 CC DOMAIN 869 913  
 CC SITE 114 115  
 CC SITE 519 521  
 CC ACT\_SITE 171 171  
 CC ACT\_SITE 212 212  
 CC ACT\_SITE 386 386  
 CC ACT\_SITE 225 225  
 CC CARBOHYD 381 381  
 CC CARBOHYD 665 665  
 CC CARBOHYD 752 752

FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 852 852 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 118 118 S -> F (IN REF. 3).  
 FT CONFLICT 121 121 V -> A (IN REF. 3).  
 FT CONFLICT 511 511 R -> A (IN REF. 3).  
 FT CONFLICT 601 601 R -> Q (IN REF. 3).  
 SQ SEQUENCE 913 AA; 101775 MW; 21389264CAD7546C CRC64;  
 Query Match 10.7%; Score 162; DB 1; Length 913;  
 Best Local Similarity 22.3%; Pred. No. 0.00014;  
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;  
 QY 20 GSONASGRGRORRMHNVNVOGCGGCGATCSDYNG--CLSCPKRPFPALEIRGMQIGVCL 77  
 DB 677 GHYADK-KRCRCAPN-----CESCFGSHGDCQMSCKYGYFL-----NEETNSCV 721  
 QY 78 SSCPSG-YGYTRYPDINKTKKADCTPCNKNPCTCKSGFYHLGKCLDNCPEGLAN 136  
 DB 722 THCPDGSYQDTKK---NLCKKSENCKTCTCFEINCTCRDGLSLQGSRCVSCEBGRYFN 778  
 QY 137 NHTME-----CVS-----IVHCEVSEW-----NPWSPCTKK 162  
 DB 779 GDCPCPCPCATCAGAGAGCINCTEGYFMEDGRCVQSCSIYTFDHSSENGYKSCCK 838  
 QY 163 GKTC-----GPKRGTE-----TRVREIIQHPSSAKGNLCPTNETRKTCTV 201  
 DB 839 DISLCITGNGFPCNCTSCPSGYLLDLGMCQGAICKDATBSWAEGGFCMLVKKNLCC-- 896  
 QY 202 QRKKCKQK 208  
 DB 897 QRKVLQK 903  
 RESULT 7  
 ID WIP1 HUMAN STANDARD; PRT; 379 AA.  
 AC Q9Y5W5; Q8WVG4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DB Wnt inhibitory factor 1 precursor (WIF-1).  
 GN WIF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Hsieh J.-C., Kodjabachian L., Rebbert M.L., Rattner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 RT activities.";  
 RL Nature 398:431-436(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,



DR PROSITE; PS01186; EGF 2; 5.  
 DR PROSITE; PS50026; EGF 3; 4.  
 DR PROSITE; PS50814; WIF; 1.  
 KW Repeat; EGF-like domain; Signal; Developmental protein;  
 KW Wnt signaling pathway.  
 FT SIGNAL 1 28  
 FT CHAIN 29 374  
 FT DOMAIN 33 171  
 FT DOMAIN 172 203  
 FT DOMAIN 204 235  
 FT DOMAIN 236 267  
 FT DOMAIN 268 299  
 FT DOMAIN 300 331  
 FT DISULFID 172 181  
 FT DISULFID 177 187  
 FT DISULFID 193 195  
 FT DISULFID 204 213  
 FT DISULFID 209 219  
 FT DISULFID 225 227  
 FT DISULFID 236 245  
 FT DISULFID 241 251  
 FT DISULFID 257 259  
 FT DISULFID 268 277  
 FT DISULFID 273 283  
 FT DISULFID 289 291  
 FT DISULFID 300 309  
 FT DISULFID 305 315  
 FT DISULFID 321 323  
 FT CARBOHYD 83 83  
 FT CARBOHYD 240 240  
 SQ SEQUENCE 374 AA; E26F973B0F00ACF8 CRC64;  
 Query Match 10.4%; Score 157.5; DB 1; Length 374;  
 Best Local Similarity 24.3%; Pred. No. 0.00013;  
 Matches 57; Conservative 22; Mismatches 85; Indels 71; Gaps 12;  
 QY 41 CQGGCAT---CSDYNGC-----LSCKPRLFFALERIGMKQIGVC-----LSSC 80  
 DB 177 CTGGCGGFCNDRHVCCEPDGFGPHCEKALCMPCMG-----GLCVTPGLGIC 227  
 QY 81 PSQYTRYVDINKTKKADCTCNKNFCTCKSGFYHLGKGLDNCPEGLBANNHTM 140  
 DB 228 PPGYGINCDKVNCTTHC-LANGTGP-----YPGKCI--CPSGYEGE----- 266  
 QY 141 ECVSIVHCVSEWNPSWPCTKKTCGFKGTETVRBIIQHPSAKGNLCPPTNETRKT 200  
 DB 267 -----QCSTSKCQ--QPCRNKGKCSG---KNKCKSGYQGLCKPVCPCSCGAHGT 315  
 QY 201 VQRKKQCKGRGKGRERK-----KPNKGSKEAIPDSKSLESKBIPE 246  
 DB 316 IEPNKCQCKE-GWNGRYCNKKYGSNLMNLRPTGSRNRQHTPSKRTEDQALPE 369  
 RESULT 9  
 PAC4 RAT  
 ID PAC4 RAT STANDARD; PRT; 937 AA.  
 AC Q63415;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)  
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like propeptide  
 DE convertase 4) (SPC4).  
 GN PACE4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Pituitary;  
 RX MEDLINE=94349873; PubMed=8070361;  
 RA Johnson R.C., Darlington D.N., Hand T.A., Bloomquist B.T., Mains R.E.;

RT "PACE4: a subtilisin-like endoprotease prevalent in the anterior  
 RT pituitary and regulated by thyroid status."; Endocrinology 135:1178-1185(1994).  
 RL Endocrinology 135:1178-1185(1994).  
 CC -!- FUNCTION: Likely to represent an endoprotease activity within the  
 CC constitutive secretory pathway, with unique restricted  
 CC distribution in both neuroendocrine and non-neuroendocrine tissues  
 CC and capable of cleavage at the RX(K/R)R consensus motif.  
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys.  
 CC -!- COFACTOR: Calcium (potential).  
 CC -!- TISSUE SPECIFICITY: High expression in the anterior pituitary and  
 CC in several brain regions, the atrium, and the ventricle.  
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone  
 CC assisting the folding of the zymogen within the endoplasmic  
 CC reticulum.  
 CC -!- SIMILARITY: Belongs to peptidase family S8.  
 CC -!- SIMILARITY: Contains 1 homo B/P domain.  
 CC -!- SIMILARITY: Contains 1 PLAC domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L31894; AAA61987.1; -.  
 DR F1R; I53282; I53282.  
 DR HSSP; Q99405; 1MPT.  
 DR MEROPS; S08.075; -.  
 DR InterPro; IPR006212; Furin repeat.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR002884; Peptidase\_S8B.  
 DR InterPro; IPR009020; Protease\_inhib.  
 DR Pfam; PF01483; P\_protease; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR ProDom; PD000717; P\_domain; 1.  
 DR SMART; SM00261; FU\_5.  
 DR PROSITE; PS00900; PLAC; 1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal; Calcium;  
 KW Cleavage on pair of basic residues; Repeat.  
 FT SIGNAL 1 45  
 FT PROPEP 46 132  
 FT CHAIN 133 937  
 FT DOMAIN 133 454  
 FT DOMAIN 477 615  
 FT DOMAIN 680 898  
 FT DOMAIN 899 937  
 FT SITE 132 133  
 FT SITE 534 536  
 FT ACT\_SITE 186 186  
 FT ACT\_SITE 227 227  
 FT ACT\_SITE 401 401  
 FT CARBOHYD 240 240  
 FT CARBOHYD 882 882  
 FT CARBOHYD 900 900  
 SQ SEQUENCE 937 AA; 104053 MW; F3865557C33705C8 CRC64;  
 Query Match 10.4%; Score 156.5; DB 1; Length 937;  
 Best Local Similarity 23.6%; Pred. No. 0.00036;  
 Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11;  
 QY 18 YIGSQNASGRQRMRHNPVSGCGGCACTCDYN--GCLSKPRLFFALERIGMKQIGV 75  
 DB 705 YFGDTAARRCR-----CHNGCETCTGRSPQCLSCR-RGPI-----HHQSTNT 747

QY 76 CLSSPCSGYCTRYPDINKCTKACDCTCNK-NFCTCKSGFYHLGKCLDNCPEGLE 134  
 Db 748 CTVLCPAGLYADESRL--CLRHPSCKQKCVCKEFTVCKEGLARGSCIPDCPQTY 805  
 QY 135 ANNHTEMC-----VSIVHCEVS-----BNWSPCTYKKGKTCGFKRGTTET 175  
 Db 806 FDESLIRGECCHTCTCVGSRBECIHCAKSPHQDKKVPACGE-----GF----- 853  
 QY 176 VREIHPKSAKGNLCPPTNTRKCTVQKCKQKGER 211  
 Db 854 -----YPEPMGL--PHKVCRRCDENCLSCBSSR 881

RESULT 10  
 FSPO\_XENLA STANDARD; PRT: 803 AA.  
 AC P35447;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE F-spondin precursor.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93376785; PubMed=8367492;  
 RA Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;  
 RT "Ectopic neural expression of a floor plate marker in frog embryos  
 injected with the midline transcription factor Pintallavis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).  
 CC -1- FUNCTION: Promotes the attachment of spinal cord and sensory  
 CC neuron cells and the outgrowth of neurites in vitro. May  
 CC contribute to the growth and guidance of axons in both the spinal  
 CC cord and the PNS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.  
 CC -1- SIMILARITY: Contains 6 TSP type-1 domains.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----  
 CC EMBL; L09123; AAA19105.1; -;  
 CC PIR; A47723; A47723.  
 CC InterPro; IPR002861; Reeler.  
 CC InterPro; IPR000884; TSP1.  
 CC Pfam; PF02014; Reeler; 1.  
 CC Pfam; PF00090; tsp1; 6.  
 CC DR SMART; SM00209; TSP1; 6.  
 CC DR PROSITE; PS00092; TSP1; 6.  
 CC Glycoprotein; Signal; Repeat; Cell adhesion.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 803 P-SPONDIN.  
 FT DOMAIN 436 489 TSP TYPE-1 1.  
 FT DOMAIN 495 549 TSP TYPE-1 2.  
 FT DOMAIN 552 605 TSP TYPE-1 3.  
 FT DOMAIN 608 662 TSP TYPE-1 4.  
 FT DOMAIN 664 717 TSP TYPE-1 5.  
 FT DOMAIN 750 802 TSP TYPE-1 6.  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 803 AA; 90702 MW; DJA54E329548AED9 CRC64;

Query Match 10.3%; Score 156; DB 1; Length 803;  
 Best Local Similarity 24.6%; Pred. No. 0.00034;

Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;  
 QY 37 VSQGGCGGATCSDYNGCLSCFPLFPALERIGMKQIGVCLSSCPG--GYGTRYPDINK 94  
 Db 544 VNEBEPSSCIVTWAEWBECS-----ATCRMGKGRHMIOTTPADGSMCKADTTTEVEK 598  
 QY 95 C--TKCK-----ADCDTCFNKNPCTKCKSGFYHLGKCLDNCPEGLEANN-- 137  
 Db 599 CMPECHTIPCVLSPWSEWSDCVTCGKTRRQR---MLKSPSELGDCNBELELQVEK 655  
 QY 138 -HTMECVSIHVCEVSEWNPSPCTKKGKTCGFKRGTTETVREIHPKSAKGNLCPPTNET 196  
 Db 656 CMLPECP--ISCELTEWSYWSBEC---NKSOG--KGHMTRMTWEPQGGAVCPETVQR 708  
 QY 197 RCTVQRKCKQKGRGKGRKGRKPKNGESKEKRAIPDSKSLSSKEIPEQRENKQ 252  
 Db 709 KKCEL--RKCK-----SSGNERRHLK-----DAREKRSSEKIKEDSDGRQ 747

RESULT 11  
 NTC2\_MOUSE STANDARD; PRT: 2470 AA.  
 ID NTC2\_MOUSE  
 AC O35516; Q06008; Q60941;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch  
 DE B).  
 GN NOTCH2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Thymus;  
 RA Hamada Y., Higuchi M., Tsujimoto Y.;  
 RT "Complete amino acid sequence and multiform transcripts encoded by a  
 RT single copy of mouse Notch2 gene.";  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 316-1518 FROM N.A.  
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
 RX MEDLINE=93178563; PubMed=8440332;  
 RA Lardelli M., Lendahl U.;  
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a  
 RT wide variety of tissues.";  
 RL Exp. Cell Res. 204:364-372(1993).  
 RN [3]  
 RP SEQUENCE OF 1765-2153 FROM N.A.  
 RX MEDLINE=97075110; PubMed=8917536;  
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,  
 RA Martin D.I.;  
 RT "Inhibition of granulocytic differentiation by mNotch1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=99396706; PubMed=10393120;  
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,  
 RA Tsujimoto Y.;  
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
 RT embryonic lethality.";  
 RL Development 126:3415-3424(1999).  
 RN [5]  
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.  
 RX MEDLINE=95333893; PubMed=7609614;  
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;  
 RT "Differential expression of Notch1 and Notch2 in developing and adult  
 RT mouse brain.";  
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).  
 RN [6]  
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
 RX MEDLINE=21523956; PubMed=11518718;



RA Saxena M. T., Schroeter E. H., Mumm J. S., Kopan R.:  
 RT "Murnine notch homologs (N1-4) undergo presenilin-dependent  
 RT proteolysis."; J. Biol. Chem. 276:40268-40273 (2001).  
 RN [7]  
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
 RX MEDLINE=21374376; PubMed=11459941;  
 RA Mizutani T., Taniguchi Y., Roki T., Hashimoto N., Honjo T.:  
 RT "Conservation of the biochemical mechanisms of signal transduction  
 RT among mammalian Notch family members."; Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).  
 RL CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands  
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.  
 CC Upon ligand activation through the released notch intracellular  
 CC domain (NICD) it forms a transcriptional activator complex with  
 CC RBP-J kappa and activates genes of the enhancer of split locus.  
 CC Affects the implementation of differentiation, proliferation and  
 CC apoptotic programs (by similarity). May play an essential role in  
 CC postimplantation development, probably in some aspect of cell  
 CC specification and/or differentiation.  
 CC CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 CC terminal fragment N(EC) which are probably linked by disulfide  
 CC bonds.  
 CC CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 CC proteolytical processing NICD is translocated to the nucleus.  
 CC CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O35516-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O35516-2; Sequence=VSP\_001405;  
 CC Note=No experimental confirmation available;  
 CC CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
 CC neuroepithelia, somites, optic vesicles and branchial arches, but  
 CC not heart.  
 CC CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
 CC the postnatal ependymal cells, and the choroid plexus throughout  
 CC embryonic and postnatal development.  
 CC CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 CC which is proteolytically cleaved by a furin-like convertase in the  
 CC trans-Golgi network before it reaches the plasma membrane to yield  
 CC an active, ligand-accessible form. Cleavage results in a C-  
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
 CC (TACE) to yield a membrane-associated intermediate fragment called  
 CC notch extracellular truncation (NEXT). This fragment is then  
 CC cleaved by presenilin dependent gamma-secretase to release a  
 CC notch-derived peptide containing the intracellular domain (NICD)  
 CC from the membrane.  
 CC CC -1- PTM: Phosphorylated.  
 CC CC -1- SIMILARITY: Belongs to the NOTCH family.  
 CC CC -1- SIMILARITY: Contains 35 EGF-like domains.  
 CC CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.  
 CC CC -1- SIMILARITY: Contains 6 ANK repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: D32210; BAA22094.1; -;  
 CC DR EMBL: X68279; CAA48340.1; -;  
 CC DR EMBL: U31881; AAC52924.1; -;  
 CC DR F01: A49175; A49175.  
 CC DR HSSP: P16109; IFSB.  
 CC DR MGD: MGI:97364; Notch2.  
 CC DR GO: GO:0005887; C:integral to plasma membrane; IC.  
 CC DR GO: GO:0005515; P:protein binding; IPI.  
 CC DR GO: GO:0002011; P:morphogenesis of an epithelial sheet; IMP.  
 CC DR GO: GO:0007219; P:N signaling pathway; IC.

DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl\_s.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001891; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_II.  
 DR InterPro: IPR006209; EGF\_like.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR008297; Notch.  
 DR InterPro: IPR008000; Notch\_dom.  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF00008; EGF; 34.  
 DR Pfam: PF00066; notch; 2.  
 DR PIRSF: PIRSF002279; Notch; 1.  
 DR PRINTS: PRO0010; EGFBL00D.  
 DR PRINTS: PRO0011; EGFFLAMIN.  
 DR PRINTS: PRO1452; NOTCH.  
 DR SMART: SM00248; ANK; 6.  
 DR SMART: SM00179; EGF\_CA; 23.  
 DR SMART: SM00004; NL; 3.  
 DR PROSITE: PS50297; ANK\_REPEAT; 1.  
 DR PROSITE: PS50098; ANK\_REPEAT; 4.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE: PS00022; EGF\_1; 33.  
 DR PROSITE: PS01186; EGF\_2; 27.  
 DR PROSITE: PS50026; EGF\_3; 35.  
 DR PROSITE: PS01187; EGF\_CA; 22.  
 DR Receptor; Transcription regulation; Activator; Differentiation;  
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation;  
 DR Alternative splicing.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
 FT CHAIN 1665 2470 NOTCH EXTRACELLULAR TRUNCATION.  
 FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.  
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1678 1698 POTENTIAL.  
 FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 26 63 EGF-LIKE 1.  
 FT DOMAIN 64 102 EGF-LIKE 2.  
 FT DOMAIN 105 143 EGF-LIKE 3.  
 FT DOMAIN 144 180 EGF-LIKE 4.  
 FT DOMAIN 182 219 EGF-LIKE 5.  
 FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).  
 FT DOMAIN 258 294 EGF-LIKE 7.  
 FT DOMAIN 296 334 EGF-LIKE 8.  
 FT DOMAIN 336 372 EGF-LIKE 9.  
 FT DOMAIN 373 411 EGF-LIKE 10.  
 FT DOMAIN 413 452 EGF-LIKE 11.  
 FT DOMAIN 454 490 EGF-LIKE 12.  
 FT DOMAIN 492 528 EGF-LIKE 13.  
 FT DOMAIN 530 566 EGF-LIKE 14.  
 FT DOMAIN 568 603 EGF-LIKE 15.  
 FT DOMAIN 605 641 EGF-LIKE 16.  
 FT DOMAIN 643 678 EGF-LIKE 17.  
 FT DOMAIN 680 716 EGF-LIKE 18.  
 FT DOMAIN 718 753 EGF-LIKE 19.  
 FT DOMAIN 755 791 EGF-LIKE 20.  
 FT DOMAIN 793 829 EGF-LIKE 21.  
 FT DOMAIN 831 869 EGF-LIKE 22.  
 FT DOMAIN 871 907 EGF-LIKE 23.  
 FT DOMAIN 909 945 EGF-LIKE 24.  
 FT DOMAIN 947 983 EGF-LIKE 25.  
 FT DOMAIN 985 1021 EGF-LIKE 26.  
 FT DOMAIN 1023 1059 EGF-LIKE 27.  
 FT DOMAIN 1061 1097 EGF-LIKE 28.  
 FT DOMAIN 1099 1145 EGF-LIKE 29.  
 FT DOMAIN 1147 1183 EGF-LIKE 30.  
 FT DOMAIN 1185 1221 EGF-LIKE 31.  
 FT DOMAIN 1223 1260 EGF-LIKE 32.  
 FT DOMAIN 1262 1300 EGF-LIKE 33.  
 FT DOMAIN 1302 1345 EGF-LIKE 34.  
 FT DOMAIN 1372 1410 EGF-LIKE 35.  
 FT REPEAT 1418 1454 LIN/NOTCH 1.



```

FT CARBOHYD 757 757 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 765 765 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 914 914 N-LINKED (GLNAC. . .) (POTENTIAL).
FT MOD RBS 1166 1166 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 985 986 AD -> PY (IN REF. 3).
SQ SEQUENCE 1370 AA; 155395 MW; A5946897A41CB145 CRC64;

Query Match 9.9%; Score 149; DB 1; Length 1370;
Best Local Similarity 26.7%; Pred. No. 0.0018;
Matches 47; Conservative 17; Mismatches 46; Indels 66; Gaps 11;

QY 34 HPNVSGCGGATCSDYNGCLSSCLSPFALERIGMKQIGVCLSSCPSSGY---GTRYP 90
Db 232 HPE---CIGSCHTPDNTTCVACRHHY-----XGVCVPACPPGTTRFEGWR-- 275
QY 91 DINKCTKCRADCDCFCNKNFCTK-----KSGFYHLGLKCLDNCPEGLEANN-HTMBC 142
Db 276 -----CVDKDFCANIPNAESSDSGDFVHDGECQECPSGFIRNSTQSMYC 321
QY 143 VSIHCEVSEWNPSCCTKGTGCGRTGTETVRRIIHPSA-----KGNL 189
Db 322 IP-----GKCP---KVCGP---DEEKTKTIDSVTAQMLQGCTILKGNL 361

RESULT 13
NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein).
GN XOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 36 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_s.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_I1.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.

```

```

DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFLOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_RBP_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1729 1750 POTENTIAL.
FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 57 EGF-LIKE 1.
FT DOMAIN 58 99 EGF-LIKE 2.
FT DOMAIN 102 140 EGF-LIKE 3.
FT DOMAIN 141 177 EGF-LIKE 4.
FT DOMAIN 179 215 EGF-LIKE 5.
FT DOMAIN 217 254 EGF-LIKE 6.
FT DOMAIN 256 292 EGF-LIKE 7.
FT DOMAIN 294 332 EGF-LIKE 8.
FT DOMAIN 334 370 EGF-LIKE 9.
FT DOMAIN 371 409 EGF-LIKE 10.
FT DOMAIN 411 449 EGF-LIKE 11.
FT DOMAIN 451 487 EGF-LIKE 12.
FT DOMAIN 489 525 EGF-LIKE 13.
FT DOMAIN 527 563 EGF-LIKE 14.
FT DOMAIN 565 600 EGF-LIKE 15.
FT DOMAIN 602 638 EGF-LIKE 16.
FT DOMAIN 640 675 EGF-LIKE 17.
FT DOMAIN 677 713 EGF-LIKE 18.
FT DOMAIN 715 750 EGF-LIKE 19.
FT DOMAIN 752 788 EGF-LIKE 20.
FT DOMAIN 790 826 EGF-LIKE 21.
FT DOMAIN 828 866 EGF-LIKE 22.
FT DOMAIN 868 904 EGF-LIKE 23.
FT DOMAIN 906 942 EGF-LIKE 24.
FT DOMAIN 944 980 EGF-LIKE 25.
FT DOMAIN 982 1018 EGF-LIKE 26.
FT DOMAIN 1020 1056 EGF-LIKE 27.
FT DOMAIN 1058 1094 EGF-LIKE 28.
FT DOMAIN 1096 1142 EGF-LIKE 29.
FT DOMAIN 1144 1180 EGF-LIKE 30.
FT DOMAIN 1182 1218 EGF-LIKE 31.
FT DOMAIN 1220 1254 EGF-LIKE 32.
FT DOMAIN 1266 1304 EGF-LIKE 33.
FT DOMAIN 1306 1346 EGF-LIKE 34.
FT DOMAIN 1347 1383 EGF-LIKE 35.
FT DOMAIN 1386 1424 EGF-LIKE 36.
FT REPEAT 1441 1478 LIN/NOTCH 1.
FT REPEAT 1479 1520 LIN/NOTCH 2.
FT REPEAT 1521 1560 LIN/NOTCH 3.
FT REPEAT 1876 1919 ANK 1.
FT REPEAT 1924 1953 ANK 2.
FT REPEAT 1957 1987 ANK 3.
FT REPEAT 1991 2020 ANK 4.
FT REPEAT 2024 2053 ANK 5.
FT REPEAT 2057 2086 ANK 6.
FT DISULFID 22 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.

```



development, probably in some aspect of cell specification and/or differentiation (By similarity).

-!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).

-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NiCD is translocated to the nucleus.

-!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.

-!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.

-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NiCD) from the membrane (By similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: Belongs to the NOTCH family.

-!- SIMILARITY: Contains 35 EGF-like domains.

-!- SIMILARITY: Contains 2 lin/Notch repeats.

-!- SIMILARITY: Contains 6 ANK repeats.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; M93661; AAK13558.1; --  
 PIR; A49128; A49128.  
 HSP; P00743; ICCP.  
 InterPro: IPR002110; ANK.  
 InterPro: IPR000152; Asx hydroxyl S.  
 InterPro: IPR000742; EGF 2.  
 InterPro: IPR001881; EGF Ca.  
 InterPro: IPR001436; EGF II.  
 InterPro: IPR006209; EGF like.  
 InterPro: IPR002049; Laminin\_EGF.  
 InterPro: IPR008297; Notch.  
 InterPro: IPR008000; Notch\_dom.  
 Pfam; PF00023; ank; 6.  
 Pfam; PF00008; EGF; 35.  
 Pfam; PF00066; notch; 2.  
 PIRSF; PIRSF002279; Notch; 1.  
 PRINTS; PR00010; EGFBLD.  
 PRINTS; PR00011; EGFAMININ.  
 PRINTS; PR01452; NOTCH.  
 SMART; SM00248; ANK; 6.  
 SMART; SM00179; EGF CA; 24.  
 SMART; SM00004; NL; 2.  
 PROSITE; PS50297; ANK REP REGION; 1.  
 PROSITE; PS50088; ANK\_REPEAT; 4.  
 PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 PROSITE; PS00022; EGF\_1; 34.  
 PROSITE; PS01186; EGF\_2; 26.  
 PROSITE; PS50026; EGF\_3; 35.  
 PROSITE; PS01187; EGF CA; 22.  
 Receptor; Transcription regulation; Activator; Differentiation;  
 Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 SIGNAL 1 25  
 CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.

FT	CHAIN	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
FT	CHAIN	1697	2471	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).
FT	DOMAIN	26	1677	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1678	1698	POTENTIAL.
FT	DOMAIN	1699	2471	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	26	63	EGF-LIKE 1.
FT	DOMAIN	64	102	EGF-LIKE 2.
FT	DOMAIN	105	143	EGF-LIKE 3.
FT	DOMAIN	144	180	EGF-LIKE 4.
FT	DOMAIN	182	219	EGF-LIKE 5.
FT	DOMAIN	221	258	EGF-LIKE 6.
FT	DOMAIN	260	296	EGF-LIKE 7.
FT	DOMAIN	298	336	EGF-LIKE 8.
FT	DOMAIN	338	374	EGF-LIKE 9.
FT	DOMAIN	375	413	EGF-LIKE 10.
FT	DOMAIN	415	454	EGF-LIKE 11.
FT	DOMAIN	456	492	EGF-LIKE 12.
FT	DOMAIN	494	530	EGF-LIKE 13.
FT	DOMAIN	532	568	EGF-LIKE 14.
FT	DOMAIN	570	605	EGF-LIKE 15.
FT	DOMAIN	607	643	EGF-LIKE 16.
FT	DOMAIN	645	680	EGF-LIKE 17.
FT	DOMAIN	682	718	EGF-LIKE 18.
FT	DOMAIN	720	755	EGF-LIKE 19.
FT	DOMAIN	757	793	EGF-LIKE 20.
FT	DOMAIN	795	831	EGF-LIKE 21.
FT	DOMAIN	833	871	EGF-LIKE 22.
FT	DOMAIN	873	909	EGF-LIKE 23.
FT	DOMAIN	911	947	EGF-LIKE 24.
FT	DOMAIN	949	985	EGF-LIKE 25.
FT	DOMAIN	987	1023	EGF-LIKE 26.
FT	DOMAIN	1025	1061	EGF-LIKE 27.
FT	DOMAIN	1063	1099	EGF-LIKE 28.
FT	DOMAIN	1101	1147	EGF-LIKE 29.
FT	DOMAIN	1149	1185	EGF-LIKE 30.
FT	DOMAIN	1187	1223	EGF-LIKE 31.
FT	DOMAIN	1225	1262	EGF-LIKE 32.
FT	DOMAIN	1264	1302	EGF-LIKE 33.
FT	DOMAIN	1304	1343	EGF-LIKE 34.
FT	DOMAIN	1374	1412	EGF-LIKE 35.
FT	DOMAIN	1645	1648	POLY-ALA.
FT	DOMAIN	1994	1997	POLY-SER.
FT	DOMAIN	2426	2429	POLY-GLY.
FT	DOMAIN	2446	2451	LIN/NOTCH 1.
FT	REPEAT	1420	1456	LIN/NOTCH 2.
FT	REPEAT	1503	1535	ANK 1.
FT	REPEAT	1827	1871	ANK 2.
FT	REPEAT	1876	1905	ANK 3.
FT	REPEAT	1909	1939	ANK 4.
FT	REPEAT	1943	1972	ANK 5.
FT	REPEAT	1976	2005	ANK 6.
FT	REPEAT	2009	2038	BY SIMILARITY.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	52	BY SIMILARITY.
FT	DISULFID	53	61	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	90	BY SIMILARITY.
FT	DISULFID	92	101	BY SIMILARITY.
FT	DISULFID	109	121	BY SIMILARITY.
FT	DISULFID	115	131	BY SIMILARITY.
FT	DISULFID	133	142	BY SIMILARITY.
FT	DISULFID	148	159	BY SIMILARITY.
FT	DISULFID	153	168	BY SIMILARITY.
FT	DISULFID	170	179	BY SIMILARITY.
FT	DISULFID	186	198	BY SIMILARITY.
FT	DISULFID	192	207	BY SIMILARITY.
FT	DISULFID	209	218	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	230	246	BY SIMILARITY.
FT	DISULFID	240	257	BY SIMILARITY.
FT	DISULFID	264	275	BY SIMILARITY.

```

FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.

Query Match 9.8%; Score 147.5; DB 1; Length 2471;
Best Local Similarity 24.8%; Pred. NO. 0.004;
Matches 54; Conservative 23; Mismatches 52; Indels 89; Gaps 13;

Qy 30 QRMRHNVQCCQGGCATCSYNGCLSCPKPLPALERIGMKQIGVCLSSCPGYYGTRY 89
Db 947 QTDVNECLSEPCXNG-GTCTSDVNSYTC-TCFAGFHGVC 984
Qy 90 P-DINKCTCKKADCTCFKNKFC-TKCKGFFVLHGK 124
Db 985 ENNIDECTE-----SSCFNGTGVGINSFSCLPVGTGPFCLHDINECSNPLNGST 1039
Qy 125 CLD-----NCPGLRANNHTECVSIVH-CRVSEMPNPSCTKKGKTCGFKRGTRV 176
Db 1040 CVDGLGTVRCTPLGYTGKN---CQTLVNLG-----SP-SPCKNKG-TCA----- 1079
Qy 177 REIQHPSAKGNLCPPTNETKCTVQRKCC-----QKG 209
Db 1080 ----QEKARPRCLCPGWDGAYCDVLNVSCKAALQKG 1113

RESULT 15
BLI4_CABEL STANDARD; PRT; 943 AA.
AC P51559; O44762; O44763; O44764; O44765; O44766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoprotease bli-4 precursor (EC 3.4.21.-) (Blisterase) (Blistered
cuticle protein 4).
GN BLI-4 OR KPC-4 OR K04F10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=95293228; PubMed=7774813;
RA Thacker C., Peters K.W., Grayko M., Rose A.M.;
RT "The bli-4 locus of Caenorhabditis elegans encodes structurally
distinct kex2/subtilisin-like endoproteases essential for early
development and adult morphology.";
RL Genes Dev. 9:956-971(1995).
RN [2]
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RA Latreille P., Wamsley P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN MUTAGENESIS.

```

```

RX MEDLINE=20363723; PubMed=10903434;
RA Thacker C., Grayko M., Rose A.M.;
RT "Mutational analysis of bli-4/kpc-4 reveals critical residues required
for proprotein convertase function in C. elegans.";
RL Gene 252:15-25(2000).
CC -!- FUNCTION: The kex2/subtilisin-like proteinase activity of this
enzyme is required for the normal production of adult cuticle.
There is functional redundancy between the isoforms. Vital for
embryonic and larval development.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=D; Synonyms=d;
CC IsoId=P51559-1; Sequence=Displayed;
CC Name=A; Synonyms=a;
CC IsoId=P51559-2; Sequence=VSP_005416, VSP_005419;
CC Name=B; Synonyms=b;
CC IsoId=P51559-3; Sequence=VSP_005418, VSP_005421;
CC Name=C; Synonyms=c;
CC IsoId=P51559-4; Sequence=VSP_005422, VSP_005423;
CC Name=E;
CC IsoId=P51559-5; Sequence=VSP_005417, VSP_005420;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: In larvae and adults, expressed in all
hypodermal cells, vulva and ventral nerve cords.
CC -!- DEVELOPMENTAL STAGE: Expression starts at embryo two-fold stage
through to adults.
CC -!- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 651.
CC -!- CAUTION: Ref.2 (AAB96754 and AAB96757) sequence differs from that
shown due to erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L29438; AAA98750.1; ALT_FRAME.
CC EMBL; L29439; AAA98751.1; ALT_FRAME.
CC EMBL; L29440; AAA98752.1; ALT_FRAME.
CC EMBL; AF039719; AAB96753.1; --.
CC EMBL; AF039719; AAB96754.1; ALT_SEQ.
CC EMBL; AF039719; AAB96755.1; --.
CC EMBL; AF039719; AAB96756.1; --.
CC EMBL; AF039719; AAB96757.1; ALT_SEQ.
CC MEROPS; S08.031; --.
CC WormPep; K04F10.4a; CE11728.
CC WormPep; K04F10.4b; CE11730.
CC WormPep; K04F10.4c; CE11732.
CC WormPep; K04F10.4d; CE11734.
CC WormPep; K04F10.4e; CE11736.
CC GO; GO:0016021; C: integral to membrane; NAS.
CC GO; GO:0004252; F: serine-type endopeptidase activity; NAS.
CC GO; GO:0007552; P: cuticle biosynthesis (sensu invertebrata); IMP.
CC InterPro; IPR006212; Furin repeat.
CC InterPro; IPR009030; Grow_fac_recep.
CC InterPro; IPR002029; Peptidase S8.
CC InterPro; IPR002884; Peptidase S8B.
CC InterPro; IPR009020; Protease Inhib.
CC Pfam; PF01483; P_protein; 1.
CC Pfam; PF00082; Peptidase S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC ProDom; PD000717; P_domain; 1.
CC SMART; SM00261; FU_3.
CC PROSITE; PS00136; SUBTILASE ASP; 1.
CC PROSITE; PS00137; SUBTILASE HIS; 1.
CC PROSITE; PS00138; SUBTILASE SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Calcium-binding; Zymogen;
KW Transmembrane; Signal; Alternative splicing.
FT SIGNAL 1 20

```

```
FT PROPEP 21 116 BY SIMILARITY.
FT CHAIN 117 943 ENDOPROTEASE BLI-4.
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 241 241 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 415 415 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT TRANSMEM 417 437 POTENTIAL.
FT DOMAIN 438 684 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 713 857 CYS-RICH.
FT CARBOHYD 195 195 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT VARSPLIC 658 670 VERSARSPFDLT -> ILITIALHLVNA (in isoform A).
FT VARSPLIC 658 699 /FTId=VSP_005416.
FT VARSPLIC 658 699 VERSARSPFDLTSGWKLSCDECNCGCTESSATSCPAYKH
FT L -> LCFNENSOPTBCVCLRCEPGVIINFQIEKLKST.
FT KCLIFS (in isoform E).
FT /FTId=VSP_005417.
FT VARSPLIC 658 730 VERSARSPFDLTSGWKLSCDECNCGCTESSATSCPAYKH
FT LITOTLRNKGSGFKVQKCDTYYLDGDKCM -> GDEVV
FT ERINHWETLEESSHNWEHREHKSILQELNLSRTHSPL
FT YSPTKFQPIFLIILVCIFDAIHRQFAV (in isoform B).
FT /FTId=VSP_005418.
FT VARSPLIC 671 943 Missing (in isoform A).
FT /FTId=VSP_005419.
FT VARSPLIC 700 943 Missing (in isoform E).
FT /FTId=VSP_005420.
FT VARSPLIC 731 943 Missing (in isoform B).
FT /FTId=VSP_005421.
FT VARSPLIC 780 828 ESNLVQAKCIWRKDLGCGDYGINAVGKDLCDSCCTCTAP
FT GPMSCERC -> AENFDPCAKNSESGRDTTVFVKFKPSAP
FT KDVCNLKLVLDPPFIPSLP (in isoform C).
FT /FTId=VSP_005422.
FT VARSPLIC 829 943 Missing (in isoform C).
FT /FTId=VSP_005423.
FT CONFLICT 153 153 A -> R (IN REF. 2).
FT CONFLICT 346 346 S -> R (IN REF. 2).
FT SQ SEQUENCE 943 AA; 103146 MW; 401E09B6C46AD7E CRC64;

Query Match 9.68; Score 145; DB 1; Length 943;
Best Local Similarity 25.6%; Pred. No. 0.0024;
Matches 51; Conservative 21; Mismatches 77; Indels 50; Gaps 11;

QY 31 RRMHPNVSQ-----COGSCATCSYNGCLSCPKRLFFALERIGMKQIGVCLSSCPSG 83
Db 663 RSSFPDLTSGWKLSCDECNCGCTESSATSCPAYK-HLTQTLRNKGGSGP-KCVQKDDT 720
QY 84 YGTRYPDINKTKCKADCTCFNKNFCTKCKSGFYL-----HL--GKCLDNCPEGLE 134
Db 721 Y-----YLDGDKCKMCSHCCTCTKAEVCETCPGSLILLIDVDMPHYDHGKCVESCPGLV 776
QY 135 ANRHTECVSIHVCESEWNPWSPCTKKGKTCGPGCTGTETVRELIQHPSAKGNLCPPTN 194
Db 777 ADYES-----NLVQAK-----CIWRKDLCG-----DGYINAVGKCDLDCSSC 814
QY 195 ETRKCT----VORKKCKOG 209
Db 815 ET--CTAPGPMSECKSKG 831
```

Search completed: June 29, 2004, 17:00:18  
Job time : 9.25243 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:55:24 ; Search time 34.6602 Seconds  
(without alignments)  
2476.067 Million cell updates/sec

Title: US-09-894-912A-34  
Perfect score: 1510  
Sequence: 1 MHLRLISWLFILNFMFYIG.....QQKKRKVDKQKSVSVSTVH 272

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_rvirus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1510	100.0	272	4 Q9BX4	Q9bx4 homo sapien
2	1457	96.5	292	4 Q9CK87	Q9ck87 homo sapien
3	1158	76.7	217	11 Q9CSB2	Q9csb2 mus musculus
4	1069	70.8	224	11 Q8BVW2	Q8bv2 mus musculus
5	644	42.6	265	11 Q9Z132	Q9z132 mus musculus
6	626.5	41.5	236	4 Q8N715	Q8n715 homo sapien
7	578.5	38.3	243	11 Q8BFU0	Q8bfu0 mus musculus
8	472	31.3	224	4 Q9UGB2	Q9ugb2 homo sapien
9	459.5	30.4	176	4 Q8N6X6	Q8n6x6 homo sapien
10	419.5	27.8	152	11 Q7TPX3	Q7tpx3 mus musculus
11	283.5	18.8	138	11 Q8BJ73	Q8bj73 mus musculus
12	192	12.7	1299	5 Q26489	Q26489 spodoptera
13	179.5	11.9	913	13 Q8A118	Q8ay18 rana escul
14	177	11.7	1101	5 Q964D2	Q964d2 entamoeba h
15	176.5	11.7	1376	5 Q8S2S2	Q8szs2 trisophila
16	176	11.7	1074	5 Q964D1	Q964d1 entamoeba h

17	172	11.4	915	11	Q91VK0	Q91vk0 mus musculu
18	167.5	11.1	296	11	Q35171	Q35171 mus musculu
19	167.5	11.1	826	11	Q8CF22	Q8cf22 mus musculu
20	167.5	11.1	932	11	Q62030	Q62030 mus musculu
21	163	10.8	3869	5	Q86PQ3	Q86pq3 cryptospori
22	162	10.7	913	4	Q96BP4	Q96bp4 homo sapien
23	161	10.7	440	5	O18003	O18003 caenorhabdi
24	159	10.5	503	5	Q9U018	Q9u018 giardia lam
25	157.5	10.4	803	13	O42114	O42114 brachydanio
26	156	10.3	898	5	O76822	O76822 brachioosto
27	155	10.3	808	13	O42113	O42113 brachydanio
28	153.5	10.2	1362	13	Q9PV24	Q9pv24 xenopus lae
29	151	10.0	238	5	O76510	O76510 cryptospori
30	149.5	9.9	4007	4	Q86XX4	Q86xx4 homo sapien
31	149	9.9	1371	11	Q9QVW4	Q9qv4 rattus sp.
32	148	9.8	548	5	Q9GQ45	Q9gq45 giardia lam
33	148	9.8	4010	11	Q80T14	Q80t14 mus musculu
34	147.5	9.8	802	13	Q9W770	Q9w770 gallus gall
35	146.5	9.7	365	11	Q924Y6	Q924y6 rattus norv
36	146.5	9.7	807	4	Q9HC86	Q9hcb6 homo sapien
37	145	9.6	724	4	O94862	O94862 homo sapien
38	145	9.6	807	4	Q8NCD7	Q8ncd7 homo sapien
39	144.5	9.6	660	5	Q23832	Q23832 cryptospori
40	144.5	9.6	2189	5	Q9BI05	Q9bi05 eimeria ten
41	144	9.5	807	6	Q9GLX9	Q9glx9 bos taurus
42	143.5	9.5	220	11	Q99KR2	Q99kr2 mus musculu
43	143.5	9.5	402	11	Q8K2Q8	Q8k2q8 mus musculu
44	143.5	9.5	807	11	Q8VCC9	Q8vcc9 mus musculu
45	143	9.5	435	5	Q9GQ41	Q9gq41 giardia lam

ALIGNMENTS

RESULT 1

Q9BX4	PRELIMINARY;	PRT; 272 AA.
ID Q9BX4		
AC Q9BX4		
DT 01-JUN-2001 (TRENBLrel. 17, Created)		
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)		
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE Thrombospondin.		
OS Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI_TaxID=9606;		
RN [1]		
RP SEQUENCE FROM N.A.		
RA Mao Y., Xie Y., Zhou Z., Zhao W., Wang W., Huang Y., Wang S.,		
RA Tang R., Chen X., Wu C.;		
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN [2]		
RP SEQUENCE FROM N.A.		
RC TISSUS=Placenta;		
RA Strausberg R.;		
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.		
DR EMBL; AF251057; AAK34947.1; -		
DR EMBL; BC022367; AAH22367.1; -		
DR Genew; HGNC:20866; THSD2.		
DR InterPro; IPR006212; Furin repeat.		
DR InterPro; IPR009030; Grow_fac_recep.		
DR InterPro; IPR000884; TSP1.		
DR Pfam; PF00090; tsp_1; 1.		
DR SMART; SM00261; FU; 2.		
DR SMART; SM00209; TSP1; 1.		
DR PROSITE; PS00092; TSP1; 1.		
SQ SEQUENCE 272 AA; 3028 MW; CACAC6B781189 CRC64;		

Query Match 100.0%; Score 1510; DB 4; Length 272;  
Best Local Similarity 100.0%; Pred. No. 3.5e-130; Indels 0; Gaps 0;  
Matches 272; Conservative 0; Mismatches 0;

Qy 1 MHLRLISWLFILNFMFYIGSQNASRGRQRMRHFNVSQGGCGCATCSYNGCLSKCP 60

Db 1 MHLRLISWLIPIILNFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60  
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGGYL 120  
 Db 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGGYL 120  
 QY 121 HLGKCLDNCPEGLAANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180  
 Db 121 HLGKCLDNCPEGLAANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180  
 QY 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRKRRKRPKNKGSKEALIPDSKSL 240  
 Db 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRKRRKRPKNKGSKEALIPDSKSL 240  
 QY 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 272  
 Db 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 272

## RESULT 2

Q96K87 PRELIMINARY; PRT; 292 AA.  
 AC Q96K87;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ14440.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arima M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.,  
 RA "NEO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK027346; BAB55051.1; --  
 DR InterPro: IPR006212; Furin repeat.  
 DR InterPro: IPR009030; Grow\_fac\_recep.  
 DR InterPro: IPR000884; TSP1.  
 DR Pfam: PF00090; tsp.1; 1.  
 DR SMART; SM00261; FU; 2.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3D4A6F8 CRC64;

Query Match 96.5%; Score 1457; DB 4; Length 292;  
 Best Local Similarity 99.2%; Pred. No. 2.7e-125;  
 Matches 263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MHLRLISWLIPIILNFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60  
 Db 1 MHLRLISWLIPIILNFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60  
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGGYL 120  
 Db 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGGYL 120  
 QY 121 HLGKCLDNCPEGLAANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180  
 Db 121 HLGKCLDNCPEGLAANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180  
 QY 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRKRRKRPKNKGSKEALIPDSKSL 240  
 Db 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRKRRKRPKNKGSKEALIPDSKSL 240

QY 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 272  
 Db 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 272  
 RESULT 3  
 Q9CSB2 PRELIMINARY; PRT; 217 AA.  
 ID Q9CSB2;  
 AC Q9CSB2 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE 2810459H04Rik protein (Fragment).  
 GN THSD2 OR 2810459H04RIK.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690 (2001).  
 DR EMBL; AK013366; BAB28811.1; --  
 DR MGI; MGI:1920030; Thsd2.  
 DR InterPro: IPR006212; Furin repeat.  
 DR InterPro: IPR009030; Grow\_fac\_recep.  
 DR InterPro: IPR000884; TSP1.  
 DR Pfam: PF00090; tsp.1; 1.  
 DR SMART; SM00261; FU; 2.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 FT NON\_TER 217 217  
 SQ SEQUENCE 217 AA; 24304 MW; 0D9F938E9FB3FBF7 CRC64;

Query Match 76.7%; Score 1158; DB 11; Length 217;  
 Best Local Similarity 93.1%; Pred. No. 4.5e-98;  
 Matches 202; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MHLRLISWLIPIILNFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60  
 Db 1 MHLRLISWLIPIILNFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60  
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGGYL 120  
 Db 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGGYL 120  
 QY 121 HLGKCLDNCPEGLAANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180  
 Db 121 HLGKCLDNCPEGLAANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180  
 QY 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRKRRKRPKNKGSKEALIPDSKSL 217  
 Db 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRKRRKRPKNKGSKEALIPDSKSL 217

```

RESULT 4
ID Q8BVW2 PRELIMINARY; PRT; 224 AA.
AC Q8BVW2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thrombospondin homolog.
GN THSD2 OR 2810459H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK076308; BAC36296.1; --
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR008212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR SMART; SM00261; FU; 2.
SQ SEQUENCE 224 AA; 25398 MW; BC13B083497CFEE3 CRC64;

Query Match 70.8%; Score 1069; DB 11; Length 224;
Best Local Similarity 84.7%; Pred. No. 6.5e-90;
Matches 188; Conservative 9; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMFYIGSQNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPR 60
Db 1 MHLRLISCFILNFMFYIGSQNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPR 60
Qy 61 LFFALERIGMKQIGVCLSSCPSSGYGTRYPDINKTKKADCDTCFNNKFTCKKSGFYL 120
Db 61 LFFVLERIGMKQIGVCLSSCPSSGYGTRYPDINKTKKADCDTCFNNKFTCKKSGFYL 120
Qy 121 HLKCLDNCPEGLRANNHTMECVSIHVCESEWNPSPCTKKGTCGFGCTETRVREII 180
Db 121 HLKCLDNCPEGLRANNHTMECVSIHVCESEWNPSPCTKKGTCGFGCTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKCTVQRKKCKGKGRERKRRK 222
Db 181 QHPSAKGNPVPNQRDNLYSTKKEVFKGRARKKGRKTKK 222

RESULT 5
ID Q9Z132 PRELIMINARY; PRT; 265 AA.
AC Q9Z132;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thrombospondin type 1 domain.
GN RSPONDIN OR R-SPONDIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;
RT "R-spondin, a novel thrombospondin type 1 domain gene, expressed in
RT the dorsal neural tube.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016768; BAA75640.1; --
DR MGD; MGI:2183426; Rspndin.

```

```

DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00092; TSP1; 1.
SQ SEQUENCE 265 AA; 29331 MW; FPEB8964743P5963 CRC64;

Query Match 42.6%; Score 644; DB 11; Length 265;
Best Local Similarity 46.5%; Pred. No. 5.9e-51;
Matches 119; Conservative 36; Mismatches 77; Indels 24; Gaps 6;

Qy 5 LISWLFILNFMFYIGSQNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPRLPFA 64
Db 11 VLSWTHIA-----VGSRGI-KGRQRRIISAGSQACAKGELCSVNGCLSKPRFIL 63
Qy 65 LERIGMKQIGVCLSSCPSSGYGTRYPDINKTKKAD-CDTCFNNKFTCKKSGFYHLG 123
Db 64 LERNDIRGVGVCLPSPGYPDARNPDNKNKCIKXIEHCEACFSHPCTKCBALYHLKG 123
Qy 124 KCLDNCPEGLRANNHTMECVSIHVCESEWNPSPCTKKGTCGFGCTETRVREIIQHP 183
Db 124 RYPACPEGSTAANSITWECGSPAQCESEWNPSPCTKKGTCGFGCTETRVREIIQHP 183
Qy 184 SAKGNLCPPTNETRKCTVQRKKCKGKGRKGGK-----RERKKKPNKSGSKBAIPDSKSLB 239
Db 184 GGDHTTCSDTETKCTVRRTPCPEGQRRKGGQRRRNNHHPARVNSKSK--PRSNS-- 239
Qy 240 SSKEIPBQRNKKQOK 255
Db 240 -----RRHKGQQQ 247

RESULT 6
ID Q8N7L5 PRELIMINARY; PRT; 236 AA.
AC Q8N7L5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ40906.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Oshima A., Takabaishi-Pujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Muraashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Katsumi H., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098225; BAC05263.1; --
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00092; TSP1; 1.
DR Hypothetical protein.
SQ SEQUENCE 236 AA; 25942 MW; 8D03803127EC5678 CRC64;

Query Match 41.5%; Score 626.5; DB 4; Length 236;
Best Local Similarity 48.1%; Pred. No. 2.1e-49;
Matches 111; Conservative 36; Mismatches 69; Indels 15; Gaps 3;

```







